UNIVERSIDAD DE CANTABRIA

e

IFCA
Instituto de Física de Cantabria

PROGRAMA DE DOCTORADO EN
ciencia y tecnología

TESIS DOCTORAL

BÚSQUEDA DE MATERIA OSCURA PRODUCIDA JUNTO A UN BOSÓN DE HIGGS EN EL CANAL DE DESINTEGRACIÓN A DOS BOSONES W^+W^- EN COLISIONES DE PROTONES A $\sqrt{s}$=13 TeV DE ENERGÍA DEL CENTRO DE MASA EN EL LHC CON EL EXPERIMENTO CMS

PH.D. THESIS

SEARCH FOR DARK MATTER PRODUCED IN ASSOCIATION WITH A HIGGS BOSON IN THE W^+W^- FULLY LEPTONIC DECAY CHANNEL IN PP COLLISIONS AT $\sqrt{s}$=13 TeV AT THE LHC WITH THE CMS DETECTOR

Realizada por:
Nicolò Trevisani

Dirigida por:
Dra. Alicia Calderón Tazón
Dra. Rocío Vilar Cortabitarte

Santander 2019

Acknowledgements

I want to spend some words to mention some of the many wonderful human beings I had the luck to meet during this part of my life.

First of all, I would like to sincerely thank my Ph.D. supervisors Alicia and Rocío for all the help and support during these years, and for making me feel not just as a student, but also as a colleague and as a friend since the first day.
My gratitude goes also to Celso Martínez, for giving me the possibility to do my Ph.D. in Santander, first by offering me a contract and later by choosing me as the recipient of the Spanish FPI grant that financed this research work.
Thanks to all my Ph.D. companions, Pedro, Cédric, Andrea, Bárbara and Celia, with whom I shared several hours of work, conferences, and spare time. I wish them all the luck and I hope they will achieve anything they want and deserve.
Many thanks to Jónatan Piedra, for the help with the code development, the suggestions on talks and presentations, and for organizing the most interesting outreach activity I know: *Las mañanas IFCA*.
Warm thanks to Pablo Martínez, for the always useful physics discussions, for sharing his clear view of the CMS working environment, and for the contagious enthusiasm he puts in everything he does and to Lara Lloret, for being there everytime I need to share a chocolate *palmera*, and for helping me with the Spanish summary of this work.
A great thanks to David Moya, for the coffees, the awesomely awful movies, all the social activities he takes care of organizing, and for being the first person arriving at IFCA in the morning and sometimes the last one to leave, despite being a *funcionario*.
Thanks to Gervasio Gómez, because it is always nice to speak Italian abroad, and for agreeing with me that the movies were actually just awesome.
Finally, thanks to Teresa Rodrigo, for the nice words and for always being available when I need her advice.

A special thanks to Javier Cuevas, for the classic Wednesday IFCA-UniOvi meetings, and for the illuminating discussions I had the pleasure to assist during the many conference dinners I had the luck to spend close to him.

This analysis work would not have been possible without the support of the *Latinos* working group, with whom I collaborated in the Higgs and mono-Higgs searches: thanks to Xavier, Piergiulio, Arun, Davide, Lorenzo, SangEun, Adrián, Isabel, Pei-Rong, Page, and all the others.

Words are not enough to explain how grateful I am to my second family, *Casa Sauverny*. For the unconditional hospitality during these years, for the Wednesday dinners, for the unlikely mountain hiking, for the *partite a FIFA*, and of course, for the marvelous fireplace. In particular, thanks to Andrea Massironi, for being my third thesis director, both at CERN and at home, to Raffaele Gerosa, for proving me that social life is possible also at CERN Restaurant 1, to Marco Pizzichemi, for sharing with me the faith in *Phazyo* (the Master) to Matteo Salomoni, for lending me his brand new car and for the *friendly* fights, to Max Ronzani, for the delicious *crostate* and for the guided tour of Santorso, and to Luca Cadamuro, for showing me something I didn't know about *Elio e le storie tese* all the times I see him.

Thanks also to my office mates at building 40: Jordi and Nuria, for the relaxing coffees during my short and rushing stays, and to Xuan and Enrique, who shared the office with me and Pedro during our longer winter stays.

A huge hug to my brothers in arms at the Universitario Cantabria Rugby Club. It is always an honor and a pleasure to be in the field with you, winning or losing. Special thanks to Manuel Samperio, who makes this possible day by day, to Carlos Juárez, for being a constant source of inspiration inside and outside the field, and to Dr. Julio Pérez and Dr. Jesús García, for sharing with me their wisdom in cooking and, most important, eating. I cannot forget about Sonia and her *La Rincuenca* bar, where we had some very good times dealing with fresh *barriles de cerveza*.

Thanks to Spanish cuisine for *tortilla de patatas*, *cocido montañes*, *cachopo*, and all the magnificent pieces of art it created, excluding, of course, *calimocho*: it is a simple disgrace.

Thanks to *AREA international popular group*, to *CCCP fedeli alla linea*, to *CSI consorzio suonatori indipendenti*, and to *Elio e le storie tese* for the amazing songs, from which I extracted the quotes at the beginning of each chapter.

Thanks to *mi piccolina* Cássia, for staying with me despite all my defects and obsessions, for always saying I'm a good cook, and for the beautiful moments we spend together every day.

Finally, the biggest thanks to my family: to my mother Loredana, my father Emanuele, my sister Carlotta, my aunt Simona, and my grandmother Maria. For the unconditional support during my studies, that brought me here, and continued unchanged during these years far from home.

Contents

Chapter 1

Introduction

Giocare col Mondo facendolo a pezzi
Bambini che il Sole ha ridotto già a vecchi

AREA, *Luglio, Agosto, Settembre (nero)*

The existence of Dark Matter (DM) is well established in modern physics. A large amount of cosmological evidence is supporting it, among them the observation that galaxies rotation velocities are not compatible with predictions from Newton's law of gravitation [1], the observation of gravitational lensing, with the spectacular example of the Bullet Cluster [2], and the anisotropies of the Cosmic Microwave Background (CMB) [3]. All the pieces of evidence suggest that DM interacts via the gravitational force, but do not give any indication on other possible interactions or on its particle nature if any. Several searches for DM have been implemented in the last decades, assuming it is made of particles which interact weakly with the known matter [1]. Many experiments looking for direct interaction of such particles within the detector material, or looking for excesses in the flux of cosmic particles due to the annihilation of DM particles into Standard Model (SM) particles have been put in place, together with searches for DM produced at colliders. No DM signals have been detected yet so that up to now only constraints have been put on the DM particle mass and on the interaction cross sections between DM and SM particles. This document focuses on the search for DM at colliders. Since DM particles are not expected to interact inside the detector volume, there traditionally one way to trigger the presence of DM has been to look for events with one energetic SM particle and large imbalance of momentum in the transverse plane, due to the presence of the undetectable DM particles, in the so-called *Mono-X* searches (where X is an SM particle). With the discovery of the Higgs boson in 2012 [4, 5], a new window in the DM search opened. If DM particles have mass and interact through the weak force, they are expected to couple to the Higgs boson, so that Higgs-boson-related signatures in colliders are a natural place to search for DM [6]. Invisible Higgs boson decays are the most direct way to explore possible DM-Higgs-boson couplings if such decays are kinematically allowed. This limits the possibility to exploit this search channel to the case $m_{DM} \leq m_h/2 \approx 60$ GeV, where m_{DM} and m_h are the DM particle and Higgs boson masses, respectively. On the other hand, the production of DM associated to a Higgs boson is sensitive to a much wider range of DM masses, since in this

[1] Many theories of physics beyond the Standard Model predict the existence of stable, neutral, weakly-interacting and massive particles (WIMPs) that are considered as Dark Matter candidates. In the following, we refer to such matter as Dark Matter.

case the Higgs boson mass is not kinematically limiting the DM production. Furthermore, while typically in Mono-X models the SM particle is emitted as Initial State Radiation (ISR), for Mono-Higgs signatures this possibility is strongly suppressed, due to the small couplings of the Higgs to light quarks and to loop-induced interaction with massless gluons. Direct coupling between the Higgs and the DM mediator is thus introduced, allowing to get information on the Higgs-DM vertex [6]. In this document, the results of the search for DM produced in association with a Higgs boson at the LHC by proton-proton collisions at a centre of mass energy of 13 TeV are presented. The data used have been collected by the CMS experiment during 2016, for a total integrated luminosity of 35.9 fb^{-1}. Following the recommendations of the ATLAS/CMS Dark Matter Forum [7] for the LHC Run 2 searches, two simplified models have been considered in the present work. In both cases, the considered decay of the Higgs boson is to a pair of W bosons, in the fully leptonic $\text{W}^+\text{W}^- \to \ell^+\nu\ell^-\bar{\nu}$ decay channel, where the two leptons are one electron and one muon. This document is organized as follows: in Chapter 2 a review of the main current shreds of evidence for the existence of Dark Matter, the candidates, the detection techniques, and the current limits are presented; in Chapter 3 the Standard Model and the Higgs mechanism are shortly introduced; in Chapter 4, the Large Hadron Collider and the CMS experiment are described, while in Chapter 5 the algorithms used to reconstruct physics objects from the electronic signals in CMS are illustrated; in Chapter 6 a short introduction to the event simulation methods is discussed; in Chapter 7 the simplified Mono-Higgs models inspected and their phenomenology are reviewed, together with the main backgrounds affecting the analysis; in Chapter 8 the event selection and the methods used to estimate the background contamination are explained; in Chapter 9 the methods to extract the signal are detailed, while the results are shown in Chapter 10. Finally, the conclusions of the document are presented in Chapter 11.

Chapter 2

Dark Matter

Ti guardo e non ti vedo
ti ascolto e non ti sento

CCCP, *And the Radio Plays*

A large amount of cosmological observations and measurements coming from different scales suggests the existence of invisible matter which interacts through the gravitational force, called dark matter. Such observations are based on the gravitational effects of dark matter on astrophysical objects. The most recent results from the Planck experiment [3] based on Cosmic Microwave Background (CMB) anisotropies measurements suggest that 5% of our Universe is made of ordinary baryonic matter, about 25% of dark matter, and the rest of dark energy [8]. Whether the dark matter is made of particles, and if it interacts with baryonic matter through other forces than gravitation is still unknown. The current experiments dedicated to dark matter search assume it is made of weakly-interacting massive particles (WIMPs), and are focused on the study of its nature and its interaction with baryonic matter. Many books and papers describing in detail the pieces of evidence for the existence of dark matter have been written. In this chapter, a brief review of the most compelling and well-known ones is presented, followed by a list of the main particle candidates for DM and a short description of the detection techniques currently employed [9].

2.1 Evidence for Dark Matter

Velocity dispersion of the Coma cluster The first hint of dark matter came from the observation of the Coma cluster. In 1933 Fritz Zwicky used the virial theorem to infer the mass of the cluster given the observed velocity dispersion of the galaxies it contains, obtaining a value around 400 times larger than the one expected considering only the presence of visible matter. This evidence lead to the introduction of the idea of invisible matter to explain the observations.

Rotation curves of galaxies One of the most convincing and direct evidence for dark matter on galactic scales comes from the observations of the rotation curves of galaxies, namely the graph of circular velocities of stars and gas as a function of their distance from

the galactic centre. In Newtonian dynamics the circular velocity $v(r)$ is expected to be:

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{2.1}$$

where r is the distance from the galactic centre, G is the gravitational cosntant, $M(r) = 4\pi \int \rho(r) r^2 dr$, and ρ(r) is the mass density profile. Considering that the large part of the stars is concentrated at the centre of the galaxy, one would expect the circular velocity to decrease as $\frac{1}{\sqrt{r}}$ at large radius. The observations, pioneered by the work of Vera Rubin [10], are in contrast with the expectations and show that $v(r)$ is approximately constant, which can be explained by the existence of a halo of dark matter with $\rho_{halo} \propto 1/r^2$. Figure 2.1 shows the observed rotation velocity for the NGC 6503 galaxy compared to the expectations assuming a spherical halo of dark matter and no coupling between dark matter and visible matter. A good agreement is observed between data and expectations, which show that the galaxy centre is mainly composed of visible matter, while the dark matter halo dominates at large radii.

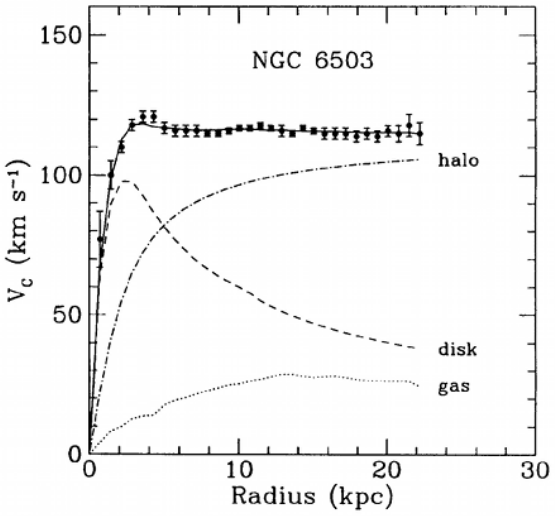


Figure 2.1: Rotation curve of NGC 6503. The dotted, dashed and dash-dotted lines are the contributions of gas, disk and dark matter, respectively [11].

Gravitational lensing Another evidence of the existence of dark matter, which arises both at the scale of galaxies and of clusters of galaxies, is gravitational lensing. Gravitational lensing is a phenomenon due to the modification of space-time caused by the presence of a large number of masses and is predicted by general relativity. The path of the photons coming from astronomical objects gets modified, leading to a distortion of their image. This allows getting an estimate of the mass of the object generating the lensing, depending on the intensity of the distortion [12, 13, 14]. The effect of lensing can be particularly evident, generating multiple images, arcs, or even Einstein rings: these are examples of strong gravitational lensing. The mass of the lens can in this case directly be inferred by the magnitude of the distortion, as for example for the Abell 370 cluster, for which a ratio between the total mass and the luminous mass of about 300 has been measured. In Figure 2.2 an image of this galaxy cluster is shown, where giant arcs can be appreciated. On the other hand, weak lensing can be used to obtain information on the

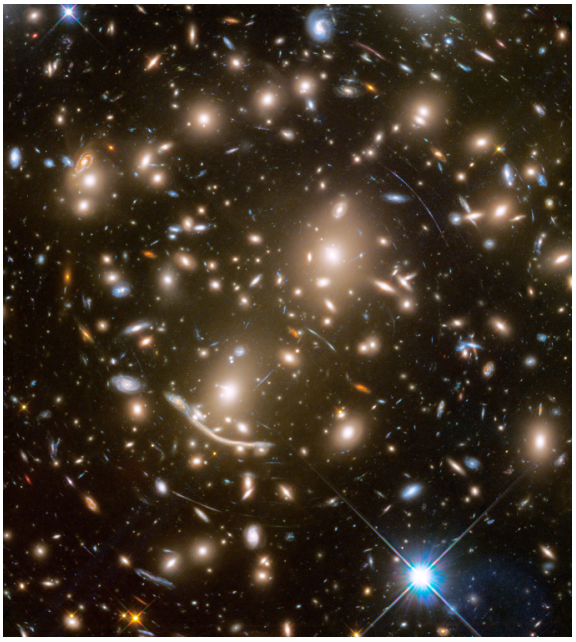


Figure 2.2: Image of the Abell 370 galaxy cluster taken by the NASA/ESA Hubble Space Telescope. Spectacular giant arcs due to strong gravitational lensing can be clearly seen [15].

dark matter distribution by looking at the effect that several galaxies or cluster of galaxies have on deep astronomical sources. By statistically analyzing the observed distortions, and with the support of strong lensing, several details of the dark matter properties can be inferred, as for example the mass profile of the dark matter halos. One well-known result obtained in this way has to deal with the Bullet Cluster. As shown in Figure 2.3, the mass profile reconstructed by weak and strong lensing shows two substructures that are offset with respect to the baryon distribution observed in X-rays by the Chandra experiment. This suggests that the two clusters are colliding, and while the baryonic matter is slowed down, dark matter is interacting much weaker and decouples. In this way, quantitative limits on dark matter self-interaction can be set [16]. Furthermore, while rotation velocity measurements can be explained also by modifying Newton's law [17], without introducing dark matter, this is much more difficult in the case of gravitational lensing, for which the assumption of the existence of dark matter appears to be more natural.

Cosmic Microwave Background A precise quantitative estimation of the amount of DM in the Universe is provided by the study of anisotropies in the CMB spectrum. The CMB is the radiation produced by the propagation of photons generated in the early Universe. It is known to be isotropic at the 10^{-5} level, and follows the spectrum of a black body corresponding to a temperature of T = 2.726 K. Through the study of its anisotropies it is possible to get precise information on the parameters of the cosmological model. This is possible since the structure of today Universe depends on the fluctuations of visible and invisible mass densities before inflation, which directly reflects on the CMB spectrum. The observed temperature anisotropies are usually expanded in spherical harmonics as:

$$\frac{\delta T}{T}(\theta,\phi) = \sum_{\ell=2}^{+\infty} \sum_{m=-\ell}^{+\ell} a_{\ell m} Y_{\ell m}(\theta,\phi) \tag{2.2}$$

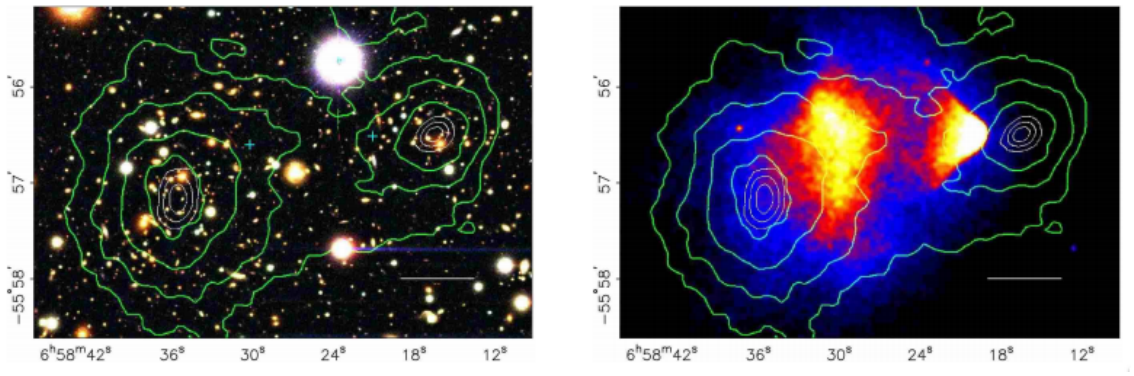


Figure 2.3: Images of the Bullet Cluster, taken by the Magellan telescope in the visible (left) and by Chandra in the X-rays, where the distribution of plasma is more evident (right). In both cases, overplotted are the contours of the spatial distribution of mass, from gravitational lensing. It is evident that most of the matter resides in a location different from the plasma.

where the information about the anisotropies is found in the coefficients $a_{\ell m}$ of the expansion, and is usually plotted as:

$$D_\ell = \frac{\ell(\ell+1)C_\ell}{2\pi} = \sum_{m=-\ell}^{+\ell} |a_{\ell m}|^2 \tag{2.3}$$

A fit to the observed data is used to extract the values of the coefficients, which in turn allow putting constraints on the parameters of the cosmological model. Among them, the baryonic matter and dark matter abundances in the Universe. The most recent and precise values are given by the Planck experiment (see Figure 2.4) [3]:

$$\Omega_b h^2 = 0.02226 \pm 0.00023 \quad \Omega_{DM} h^2 = 0.1186 \pm 0.0020 \tag{2.4}$$

These values translate in a fraction of 4.8% of ordinary matter, and 25.8% of dark matter.

2.2 Dark Matter Candidates

2.2.1 Ordinary matter

The original meaning of dark matter indicated any kind of matter which is invisible and could in principle explain the observations listed in Section 2.1. This did not exclude baryonic matter in the form of compact objects, like planets, brown dwarfs, red dwarfs, white dwarfs, neutron stars and black holes, which are much less luminous than ordinary stars. On the other hand, when the idea of particle dark matter was introduced, one natural candidate was the neutrino, which had the "undisputed virtue of being known to exist" [18]. Ordinary matter has been excluded as a candidate for dark matter, but it is still interesting to understand why.

Massive Astrophysical Compact Halo Objects Massive Astrophysical Compact Halo Objects (MACHOs) are natural candidates to constitute dark matter if it is made up of baryonic matter. Even if not directly visible, MACHOs can be detected through the

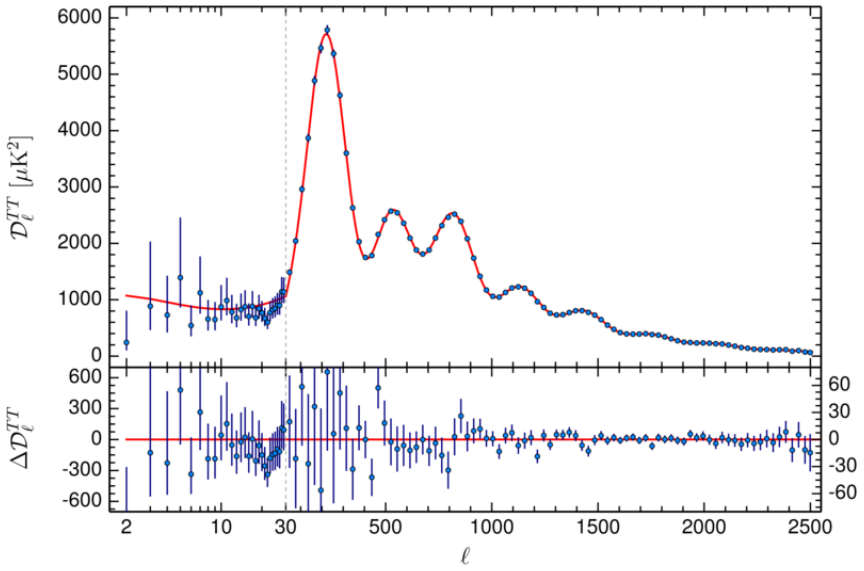


Figure 2.4: Power spectrum of the CMB anisotropies obtained by the Planck experiment 2015 data. The data points (blue dots) are fitted to the spectrum predicted by a baseline Λ_{CDM} model (red line). The lower panel shows the residuals with respect to the best-fit spectrum [3].

gravitational microlensing induced by their mass. Dedicated experiments, as EROS [19] and MACHO [20], have looked for signs of the presence of MACHOs in our galaxy. In Figure 2.5, the combined results of the two experiments are shown and exclude that objects with mass M 0.6×10^{-7} $\mathrm{M}_\odot < \mathrm{M} < 15\mathrm{M}_\odot$, where $\mathrm{M}_\odot$ is the solar mass, can constitute a large fraction of dark matter in the galaxy halo. These results, together with the limits on the baryonic matter abundance set by cosmic microwave background anisotropies, exclude that MACHOs can significantly contribute to dark matter.

Neutrinos Considering that neutrinos interact only weakly and gravitationally, the hypothesis that they form part of the dark matter budget is in principle more than legit. The total relic density of neutrinos can be written as:

$$\Omega_\nu h^2 = \sum_{i=1}^{3} \frac{m_i}{93 \text{ eV}} \tag{2.5}$$

where m_i is the mass of the i-th neutrino. This means that neutrinos, to account for all the dark matter relic density should have a $m_{tot} = \sum_{i=1}^{3} m_i \sim 11$ eV. Direct measurements of the mass of the electron antineutrino set an upper limit of $\mathrm{m}_\nu < 2.05$ eV. The three neutrinos must have very similar masses, as required by the observation of neutrino oscillations, so that neutrinos are not abundant enough to be the main component of dark matter. Furthermore, neutrinos are highly relativistic: relativistic dark matter would have favoured a Universe in which small structures formed after large structures, while several observations seem to indicate that small structures formed first [21, 22].

2.2.2 WIMPs

One very appealing hypothesis regarding the particle nature of dark matter is that it is made up of weakly interacting massive particles (WIMPs) with a mass of the order of

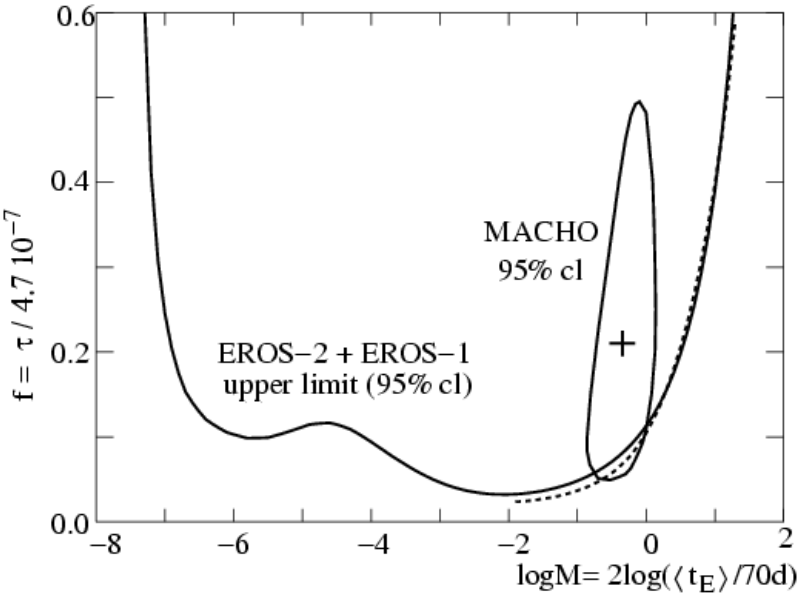


Figure 2.5: Excluded fraction at 95% confidence level (CL) of DM halo constituted by MACHOs, f, as a function of the MACHO mass, M, for the combined analysis of the EROS surveys [19].

the electroweak symmetry breaking scale (10 GeV - 1 TeV). An important reason for the popularity of this assumption is the fact that WIMPs naturally have the relic abundance of the right order of magnitude predicted by observations. Moreover, since they are expected to have an interaction strength similar to the Standard Model electroweak particles, they are of particular interest also from an observational point of view. A detailed discussion regarding the WIMP relic abundance calculations can be found in [9, 23]. In this section, a summary of the key points is presented. In the early Universe, DM and SM particles were thought to be in thermal equilibrium with each other, with production and annihilation of WIMPs pairs in particle-antiparticle pairs, as:

$$\chi\bar{\chi} \longleftrightarrow e^+e^-, \mu^+\mu^-, q\bar{q}, W^+W^-, ZZ, hh, ... \tag{2.6}$$

With a temperature much higher than the WIMP mass, $T \gg m_\chi$, particle-antiparticle collisions have enough energy to produce WIMPs efficiently, and the production of DM particles is in equilibrium with DM annihilation. In these conditions, the common rate of annihilation is:

$$\Gamma_{\text{ann}} = \langle \sigma_{\text{ann}} v \rangle n_{\text{eq}} \tag{2.7}$$

where σ_{ann} is the WIMP annihilation cross-section, v is the relative velocity of the annihilating WIMP, n_{eq} is the WIMP number density at equilibrium, and the brackets indicate the average over the WIMP thermal distribution. Due to the expansion of the Universe, the temperature of the particle plasma became smaller than the WIMP mass. Only particle-antiparticle collisions in the tail of the Boltzmann distribution had enough energy to produce WIMP pairs, so the number of WIMPs produced decreased as $e^{-m_{\chi/T}}$. The other consequence of Universe expansion is that also the number density of particles n became smaller, making both the WIMP production and annihilation rate decrease. In the end, the WIMP annihilation rate became smaller than the expansion rate of the Universe

$H(t)$ [1], or equivalently the mean free path for WIMP-producing collisions became longer than the radius of the Universe, and the WIMP production stopped (decoupling). This translates in the fact that the number of WIMPs in co-moving volume remained constant, or in other words their number density decreased with the volume V, like $\sim \frac{1}{\mathrm{V}}$. As said, the popularity of the WIMP model is given by its capability of predicting the current density of DM in the Universe. To compute the current density of WIMPs, the rate equation for the WIMP number density n and the law of entropy conservation can be exploited:

$$\begin{cases} \frac{dn}{dt} = -3Hn - \langle\sigma_{\mathrm{ann}}v\rangle(n^2 - n_{\mathrm{eq}}^2) \\ \frac{ds}{dt} = -3Hs \end{cases} \tag{2.8}$$

where t denotes time, s the entropy density, H the Hubble parameter, and where the right-hand side of the first equation in 2.8 takes into account the dependence of the WIMP number density on the expansion of the Universe and on the change in number density due to annihilation and inverse annihilation. By defining the co-moving number density $\mathrm{Y} = \frac{n}{s}$ and by using the photon temperature T instead of time as an independent variable (photon temperature decreases monotonically with time) through the definition of $\mathrm{x} = \frac{m_\chi}{T}$, one can re-write the first equation in 2.8 as:

$$\frac{dY}{dx} = \frac{1}{3H}\frac{ds}{dx}\langle\sigma_{\mathrm{ann}}v\rangle(n^2 - n_{\mathrm{eq}}^2) \tag{2.9}$$

which is the most common way of expressing the evolution of WIMP abundance. Equation 2.9 cannot be solved analytically; numerical solutions are shown in Figure 2.6. There it can be seen that while $\mathrm{T} \geq \mathrm{m}_\chi$, thermal equilibrium is maintained and the WIMP abundance decreases as a Boltzmann exponential. Without expansion, the trend would be followed as shown by the dashed line. Since the Universe is expanding, not just the WIMP production is suppressed, but also its annihilation, as the WIMP collision rate decreases. Once the annihilation rate cannot keep up with the expansion, thermal equilibrium is lost and the co-moving number density remains approximately constant. This occurs when:

$$\Gamma_{\mathrm{ann}} = n\langle\sigma_{\mathrm{ann}}v\rangle \sim H \tag{2.10}$$

By imposing this condition in the definition of the current DM density and remembering the value measured by Planck, we obtain:

$$\Omega_{\mathrm{DM}}h^2 = \frac{\rho_{\mathrm{DM}}}{\rho_{\mathrm{c}}}h^2 = \frac{m_{DM}}{n}\rho_{\mathrm{c}}h^2 = \frac{m_{DM}}{\rho_{\mathrm{c}}}\frac{H_0}{\langle\sigma_{\mathrm{ann}}v\rangle}h^2 = 0.1186 \tag{2.11}$$

where ρ_{c} is the critical density, namely the energy density required for the Universe to be flat, and $h = H_0/100\ \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, with H_0 being the Hubble constant. Substituting now the numerical values for:

- $H_0 = 2.176 \times 10^{-18}\ \mathrm{sec}^{-1}$ [3]
- $\rho_{\mathrm{c}} = \frac{3H^2}{8\pi G} = 1.8788 \times 10^{-26}\ h^2\ \mathrm{Kg/m}^3$

[1] Defining the Universe radius at the time t as $a(t)$, the Universe expansion rate at a time t is defined as $H(t) = \frac{\dot{a}(t)}{a(t)}$. $H(t)$ is known as Hubble parameter. The current value of the Hubble parameter is indicated as Hubble constant, or H_0.

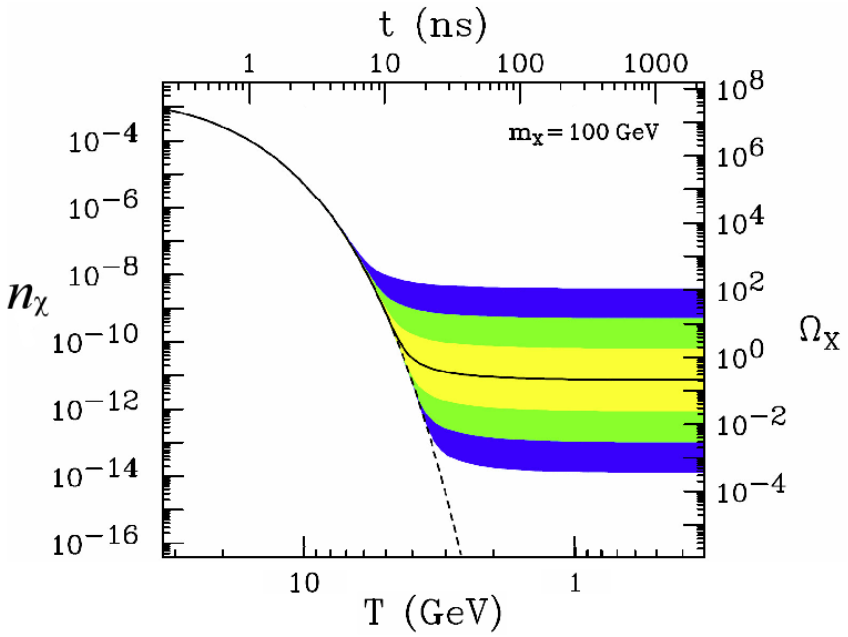


Figure 2.6: The number density n_χ and resulting thermal relic density Ω_χ for a 100 GeV WIMP as a function of temperature or time. The solid line corresponds to the solution of the Boltzmann equation which yields the correct relic density, and the coloured regions correspond to cross sections that differ 10, 100 or 1000 times from this value. The dashed line indicates the solution for a particle that remains in thermal equilibrium [24].

and assuming for the WIMP a mass of the order of the electroweak symmetry breaking scale (e.g. 100 GeV) one gets for $\langle \sigma_{\text{ann}} v \rangle$ a value of about 3×10^{-26} cm^3/sec, close to what expected for weak interactions. The fact that assuming that dark matter is made of weakly interacting massive particles with a mass close to the energy of the electroweak scale and a cross-section similar to the ones measured for electroweak processes naturally returns the correct dark matter abundance is known as "WIMP miracle".

WIMP candidates

Different extensions of the Standard Model are able to predict the existence of WIMPs. Here the two most popular ones are shortly introduced. A more exhaustive review can be found in Reference [24].

Supersymmetry The basic assumption of supersymmetry [25] (SUSY) is that for every fermion a boson must exist with the same quantum numbers, and vice-versa. If confirmed by experimental observations, supersymmetry would solve fundamental questions of particle physics, as the hierarchy problem, affecting particularly the radiative corrections to the Higgs boson mass. One consequence of this model is the introduction of several new particles with masses of the order of the 100 GeV - 1 TeV, among them electrically neutral and non-strongly interacting particles, like the superpartners of neutrino (sneutrino), graviton (gravitino), photon, Z boson, and Higgs boson (neutralinos). In particular, the lightest supersymmetric particle, typically the neutralino or the sneutrino, is considered as a natural WIMP candidate.

Universal Extra Dimension Theory Extra dimensions were originally introduced in an attempt of unifying gravitation and electromagnetism, in the Kaluza-Klein theory [26]. More recent versions of this same idea are used to produce models in which all the Standard Model particles live in five or six-dimensional space-time, in which one or two dimensions are compactified with radius R^{-1} [27]. These models are known as Universal Extra Dimension and predict the existence of new particles (Kaluza-Klein resonances) at an energy scale related to the inverse size of the extra dimension: $m_{KK} \sim R^{-1}$. The lightest and stable decay products of these resonances can be valid WIMP candidates.

2.2.3 Non-WIMP candidates

Despite its appeal, due to the capability of naturally predicting the relic abundance, the WIMP hypothesis is not the only one able to explain the particle nature of dark matter. Two interesting examples are axions and sterile neutrinos, both supported by strong physics motivations. In addition, there are models which, as in the case of WIMPs, can reproduce the correct relic density and introduce particles that may be seen in current experiments.

Axions Axions were first introduced to solve the so-called *strong-CP* problem. The problem arises due to the presence in the QCD Lagrangian of the following term:

$$\mathscr{L} \supset \bar{\Theta}\frac{g^2}{32\pi^2}G^{a\mu\nu}\widetilde{G}_{a\mu\nu} \tag{2.12}$$

where $G^{a\mu\nu}$ is the gluon field strength tensor and $\bar{\Theta}$ is related to the phase of the QCD vacuum. A value of $\bar{\Theta}$ close to 1, as would naively be expected, would lead to large CP-violating effects, causing the electric dipole moment of the neutron to be almost 10^{10} times larger than observed. This inconsistency could be solved by simply putting the correct value *by hand*, or by assuming that some new physics is causing $\bar{\Theta}$ to be so small. One very promising solution was proposed by Peccei and Quinn [28, 29], who introduced a new global U(1) symmetry that is spontaneously broken, and showed that this can dynamically drive $\bar{\Theta}$ toward zero. This new symmetry, as pointed out by Wilczeck [30] and Weinberg [31] implies the existence of a new Nambu-Goldstone boson, called axion. Initially, the mass of this new particle was thought to be at the scale of the MeV, but quickly both laboratory and astrophysical observations put strong constraints on its mass, with a current upper limit of the order of the meV. This new mass regime and a low expected interaction strength with matter make axions an interesting candidate for particle dark matter, making them stable over cosmological timescale and, if present with the correct abundance, able to explain the observations associated to the presence of dark matter in the Universe.

Sterile neutrinos In the Standard Model neutrinos are considered as massless particles, and the three leptonic numbers associated to the e, μ, and τ flavours are separately conserved. Nevertheless, the observation of neutrino oscillation implies that such particles are actually massive, and that lepton flavour violation is permitted. Different extensions of the Standard Model have been proposed to solve this issue, and one of them assumes that a fourth neutrino exists, which interacts with the other neutrino families by oscillation but is not affected by electroweak force. This new particle is indicated as the sterile neutrino and is considered as a potential dark matter candidate [32].

Super-WIMP dark matter In the super-WIMP framework the assumption is that WIMPs freeze out as usual in the early Universe, but then decay to superweakly interacting

massive particles, which form the currently existing dark matter. Super-WIMPs can be accommodated in different theories beyond the Standard Model, like SUSY. In this case, the super-WIMP is represented by the lightest supersymmetric particle (LSP), typically the gravitino, and the WIMP by the next-to-lightest supersymmetric particle (NLSP), which can be a slepton, a sneutrino or a neutralino.

WIMP-less dark matter The "WIMP miracle" is due to the fact that assuming dark matter particles interact weakly and their mass is close to the one of other weakly-interacting particles, one gets the correct dark matter abundance. Since, in general:

$$\Omega_{DM} \propto \frac{1}{\langle \sigma v \rangle} \sim \frac{m_{DM}^2}{g_{DM}^4} \tag{2.13}$$

It is possible to tune the mass and the coupling of the particle candidates in order to get the correct relic density with $(\mathrm{m}_{DM}, g_{DM}) \neq (\mathrm{m}_{\mathrm{weak}}, g_{\mathrm{weak}})$. In this case, the dark matter would not interact weakly (WIMP-less dark matter) so no gauge interactions with SM particles are expected.

2.3 Dark matter detection

Different approaches are used to detect particle dark matter. Direct detection is the most straightforward way one can think: assuming that the galaxy is full of WIMPs, many of them should pass through the Earth and interact within experiments, mainly through scattering with nuclei of the material. The possible products of dark matter annihilation, like photons, neutrinos, or charged particles, are instead the targets of indirect detection experiments. Last, but not least, colliders are potentially factories of dark matter. There the expected signature is a large imbalance of the transverse momentum, due to the fact that WIMPs are not supposed to interact in the relatively low volume of the detectors. In this section, a short review of the basic ideas used for dark matter detection and the main results are presented.

2.3.1 Direct detection

If our galaxy's halo is composed of WIMPs, then their flux on the Earth is of the order of $10^5 (100 \text{ GeV}/\mathrm{m}_\chi)$ cm^{-2}s^{-1}. The aim of direct detection experiments is to detect the small fraction of elastic interactions that would take place within specifically designed low background experiments, measuring the rate and the energies of the nuclear recoils. The energy differential event rate for a WIMP with mass m_χ and a nucleus with mass m_N, expressed in terms of counts kg^{-1} day^{-1} keV^{-1}, is given by:

$$\frac{\mathrm{d}R}{\mathrm{d}E_\mathrm{R}} = \frac{\rho_0}{m_N m_\chi} \int_{v_{min}}^{\infty} v f(v) \frac{\mathrm{d}\sigma_{WN}}{\mathrm{d}E_\mathrm{R}}(v, E_\mathrm{R}) \mathrm{d}v \tag{2.14}$$

where:

- ρ_0 is the local WIMP density. In most of the experiments, a value of 0.3 GeV cm^{-3} is considered.
- $f(v)$ is the WIMP velocity probability density function in the detector frame. To estimate it, one has to take into account both the speed of WIMPs and of the Earth. For the former, the halo is considered an isothermal sphere with density profile $\rho(\mathrm{r}) \propto \mathrm{r}^{-2}$;

this returns an isotropic, Gaussian velocity distribution. Even if this assumption has been shown to be valid only as a first approximation, for experiments which look for rates integrated over time this does not represent an issue, since by integrating over $f(v)$ the dependence on the details of the model is reduced. For the latter, one has in principle to consider three contributions: the motion of the local standard of rest (LSR), the Sun's peculiar motion with respect to the LSR, and the Earth's orbit around the Sun. For experiments not seeking for time or directional modulation, only the first contribution is relevant. The circular velocity of the LSR obtained by combining a large number of measurements is found to be $v_c = (220 \pm 20)$ km s^{-1}.

- The nucleon recoil energy E_{R} can be easily calculated in terms of the scattering angle in the centre of mass frame θ^*, thanks to the fact that the WIMP-nucleon relative speed of the order 100 km/s allows calculating the elastic scattering in the extreme non-relativistic limit:

$$E_{\mathrm{R}} = \frac{\mu_N^2 v^2 (1 - \cos\theta^*)}{m_N} \tag{2.15}$$

  where $\mu_N = \dfrac{m_\chi m_N}{(m_\chi + m_N)}$ is the WIMP-nucleus reduced mass.

- The lower limit of the integration over WIMP speed corresponds to the minimum WIMP speed which can generate a recoil of energy E_{R}:

$$v_{min} = \sqrt{\frac{m_N E_{\mathrm{R}}}{2\mu_N^2}} \tag{2.16}$$

The total event rate is then calculated by integrating the differential event rate over all the possible recoil energies:

$$R = \int_{E_{\mathrm{T}}}^{\infty} \mathrm{d}E_{\mathrm{R}} \frac{\rho_0}{m_N m_\chi} \int_{v_{min}}^{\infty} v f(v) \frac{\mathrm{d}\sigma_{WN}}{\mathrm{d}E_{\mathrm{R}}}(v, E_{\mathrm{R}}) \mathrm{d}v \tag{2.17}$$

There:

- E_{T} is the threshold energy, namely the minimum recoil energy needed to generate a signal in the detector.

- $\dfrac{\mathrm{d}\sigma_{WN}}{\mathrm{d}E_{\mathrm{R}}}(v, E_{\mathrm{R}})$ is the differential cross-section for the WIMP-nucleus elastic scattering. It contains the information on the dark matter particle physics, including the WIMP interaction properties with nuclei. It is generally expressed in terms of a spin-dependent and a spin-independent components, each one written as a cross-section and a form factor, which takes into account the dependence on the momentum transfer:

$$\frac{\mathrm{d}\sigma_{WN}}{\mathrm{d}E_{\mathrm{R}}} = \frac{m_N}{2\mu_N^2 v^2} \left(\sigma_0^{\mathrm{SI}} F_{SI}^2(E_{\mathrm{R}}) + \sigma_0^{\mathrm{SD}} F_{SD}^2(E_{\mathrm{R}}) \right) \tag{2.18}$$

  where SI and SD stand for spin-dependent and spin-independent, respectively. The results of direct-detection experiments are usually expressed in terms of the two cross-sections σ_0^{SI} and σ_0^{SD}.

Detection techniques and current limits

Direct detection experiments make use of a wide set of materials and technologies, but all of them share a set of basic characteristics, dictated by the need of detecting rare events with low energies. The first requirement is fulfilled by designing large mass detectors, to increase the interaction probability, while the second condition needs instruments with very low energy thresholds since, for WIMPs with masses between 1 GeV and 1 TeV, recoil energies from 1 KeV to 100 KeV are expected. These two characteristics make direct detection experiments very sensitive to any possible sources of backgrounds. To reduce the rate of interactions due to cosmic rays, these experiments are usually operated in deep underground laboratories, and to suppress the background coming from neutrons and photons in the laboratory environment, and from the detector components themselves, shielding techniques and high purity materials are employed. Different detection techniques are currently used in direct detection experiments, mainly ionization, scintillation, and phonons production, based on several materials and technologies. In Figure 2.7 some of the main experiments are categorized according to the WIMP detection strategy employed. The

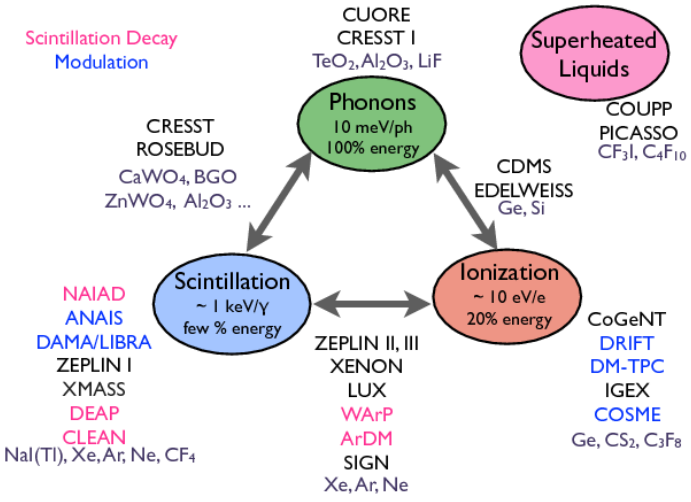


Figure 2.7: Diagram listing direct detection experiments based on their detection strategy and on the materials used [33].

results of direct detection experiments are presented as limits on WIMP-nucleon interaction cross-section (σ^{SI} and σ^{SD}) as a function of the WIMP mass. In the SD case, the limits are separated between neutron and proton, while for SI the two components are assumed to be equal. Currently, the most stringent limits are set to the spin-independent component. In Figure 2.8 the current SI limits are shown, with exclusion reaching cross sections of 10^{-46} cm^2 for WIMP masses between 10 GeV and 100 GeV. In the same plot, the projections from a small number of upcoming experiments are also shown, together with the irreducible background coming from solar and atmospheric neutrinos. The uncertainty on the rates induced by this process is the main issue for the performances of these experiments. For the spin-dependent component, experiments are usually sensitive

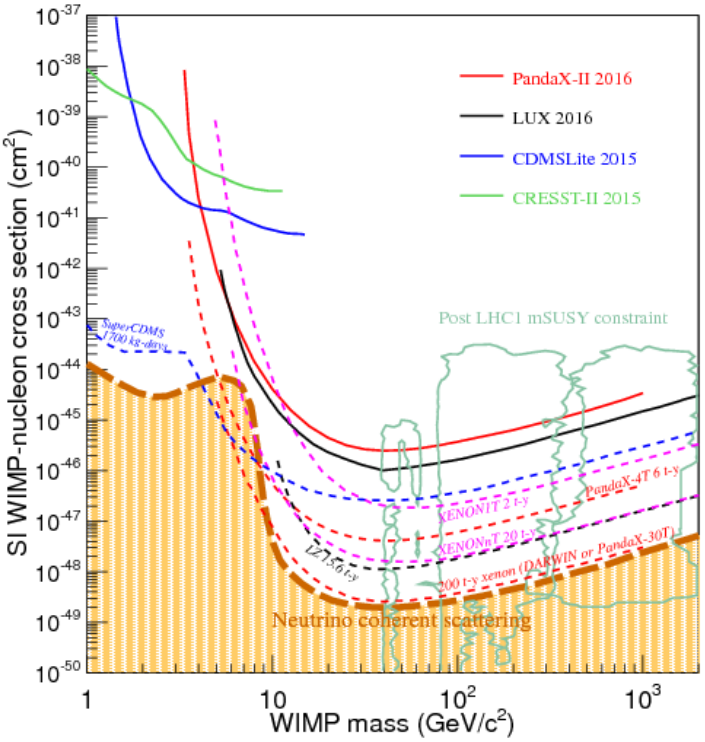


Figure 2.8: Current limits and future projections on the spin-independent WIMP-nucleons cross section from direct detection experiments [34].

either to the WIMP-neutron or the WIMP-proton interaction. For the former, the best limits are provided by LUX [35], and for the latter by PICO [36], as shown in Figure 2.9. In the plots, also the limits set by several indirect detection and collider experiments are presented. With the exception of the claim for a signal reported by the DAMA/LIBRA

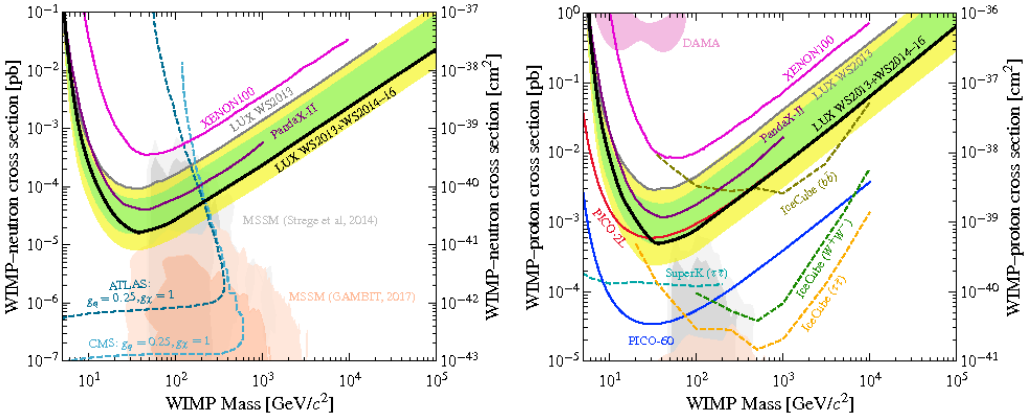


Figure 2.9: Current limits on the spin-dependent WIMP-neutron (left) and WIMP-proton (right) cross sections reported by the LUX collaboration [35]. In the case of the WIMP-proton scattering, limits from indirect detection experiments (SuperKamiokande and IceCube) are the most stringent in many regions of the parameter space.

collaboration [37], which is denied by other experiments, no solid evidence for dark matter

has been provided by direct detection experiments so far.

2.3.2 Indirect detection

Indirect detection of dark matter consists of looking for the radiation produced by dark matter annihilation or decay. This includes mainly gamma rays, neutrinos, and charged particles. The flux of this radiation is proportional to the annihilation rate, which depends on the square of the dark matter density, $\Gamma_A \propto \rho_{DM}^2$. For this reason, the best places where to look for such radiation are the regions where large dark matter densities are expected, as the galactic centre, or astrophysical objects like the Earth and the Sun, where dark matter can be captured due to the loss of energy caused by the scattering of WIMPs with the nucleons in the interiors of such objects. The indirect search for DM follows three main branches, which are listed and shortly described below:

- One of the most interesting characteristics of the indirect search for dark matter decaying to **gamma rays** is that the expected signal, in this case, is a mono-energetic line in the photon spectrum. If a signal like that, incompatible with other astrophysical gamma rays source, would be detected, it would represent clear evidence of the existence of dark matter. Gamma rays are detected through large imaging air Cherenkov telescopes, like HESS [38] and MAGIC [39], and through space telescopes, such as Fermi Gamma-ray Space Telescope [40]. In particular, this last experiment observed a line at an energy of approximately 130 GeV, giving rise to several discussions in the last years. Following analyses showed a reduced significance of the excess with respect to the original observation, and the interpretation of the signal as a piece of evidence for dark matter is now not supported even by the Fermi collaboration [41]. More recently a new excess at an energy of few GeV has been observed by Fermi, confirmed by subsequent analyses [42]. Although the excess, in this case, is firmly confirmed, its interpretation as self-annihilating DM particles is not unambiguously possible due to difficulties in the modelling of the systematic uncertainties affecting gamma rays in the galactic plane [43].

- **Neutrinos** can be secondary products of dark matter annihilation inside the Sun. Moving through the halo of the Milky Way, the Sun interacts with WIMPs, which by collisions with its nuclei can lose energy and get eventually gravitationally trapped. Over the lifetime of the Sun, an equilibrium between the ratios of WIMPs trapping and decaying is reached. The DM particles can then annihilate to several SM particles; some of them, like vector bosons and quarks, can further decay producing neutrinos which can easily escape the Sun and reach the Earth. Here they can be detected using large-volume neutrino detectors like SuperKamiokande [44] and Ice Cube [45]. These experiments aim to detect the Cherenkov radiation due to the interaction of secondary particles produced by the interaction of the neutrinos with the detector medium. No excess of events with respect to expected background has been observed yet, so limits on the velocity averaged WIMP annihilation cross section are set. These results can be translated to WIMP-nucleon scattering cross sections and compared with direct-detection experiment results. As can be seen in Figure 2.9, indirect search for dark matter through neutrinos experiments gives the most stringent results for WIMP masses above 100 GeV.

- A more difficult target is the production of **charged particles** due to the annihilation of DM pairs since a large number of effects can influence the propagation of such particles. Usually, the searches in this case focus on positrons and antiprotons,

due to the lower abundance of anti-matter with respect to ordinary matter in cosmic rays produced by standard, non-DM sources. Different experimental techniques are exploited to look for charged particles, from detectors on balloons (e.g. ATIC [46]) or in space (e.g. PAMELA [47] and AMS [48]), to large-area cosmic-ray detectors on the ground (e.g. Auger [49]). Several experiments observed an increase in the positrons fraction in the range from 10 GeV to 500 GeV, as can be seen in Figure 2.10, while standard expectations for galactic cosmic rays predict a falling spectrum. Two equivalently valid explanations have been proposed to explain this excess: dark matter annihilation and the production of cosmic rays by pulsars. To distinguish between the two hypotheses, more precise results on the positron fraction and the antiproton-to-proton ratio will be needed. A flat behaviour in the antiproton-to-proton flux at high energies has been highlighted by the most recent AMS-02 results, as shown in Figure 2.11. This is thought to be incompatible with secondary production of antiprotons from ordinary cosmic rays and of difficult explanation from pulsar origin [50].

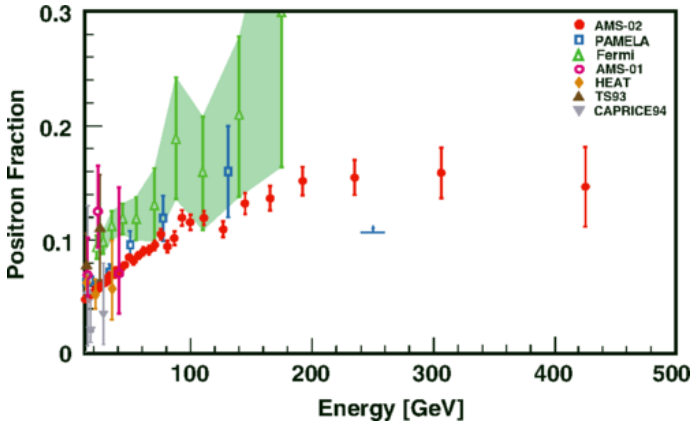


Figure 2.10: The positron fraction above 10 GeV, where it begins to increase, as measured by the AMS-02 experiment and compared to the results of other experiments [51].

2.3.3 Detection at Colliders

If DM particles can interact weakly with SM particles, they can be produced at a collider: for this reason, WIMPs are the best candidates for collider searches. Furthermore, since they are expected to present both production cross sections and masses of the order of the electroweak scale, the LHC is the ideal machine to produce them. Since WIMPs have to be stable over the Universe lifetime, they are not expected to decay in the detector. In addition, since they interact only weakly, no signals in the detectors are expected. Nevertheless, different strategies have been developed to trigger their presence at the ATLAS, CMS, and LHCb detectors located along the LHC.

- **Mediator searches**: In this case a dark matter mediator is produced by the annihilation of a pair of quarks or gluons, and decays to a pair of quarks or leptons, producing a localized excess of events in the invariant mass spectrum [52, 53, 54], or in specifically chosen angular distributions [55].

- **Mono-X searches**: The dark matter mediator, in this case, decays to a pair of WIMPs, which escape the experimental device undetected. To trigger this kind of events, the presence of a detectable SM particle in the event is required [56, 57, 58,

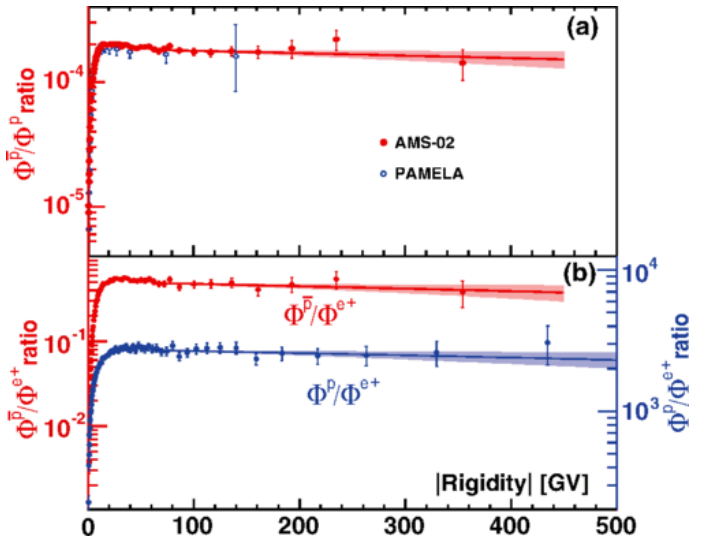


Figure 2.11: Top plot: the $(\bar{p}/p)$ flux ratio as a function of the absolute value of the rigidity (momentum per unit of charge) from 1 to 450 GeV as measured by AMS and compared with the results by PAMELA. Bottom plot: the AMS results for $(\bar{p}/e^+)$ (on the left) and for (p/e^+) (on the right) flux ratios [48].

59, 60, 61, 62, 63, 64, 65]. The fact that the WIMPs cannot be recorded creates an imbalance in the momentum measured in the transverse plane, called *missing transverse energy* (E_T^{miss}) or *missing transverse momentum* (p_T^{miss}), which is a key property of these searches.

- **Long-lived searches**: The DM mediators can have a relatively long lifetime, leading to signatures in which the SM particles come from a displaced vertex with respect to the point of protons interaction. This kind of events is hardly produced by SM processes so that long-lived searches present a small amount of background and are becoming popular at the LHC [66].

- **Invisible Higgs decays**: If the DM particles acquire mass through electroweak symmetry breaking, a coupling between DM and the Higgs boson naturally arises. If additionally, this mass is smaller than half the mass of the Higgs boson, the decay channel $h \rightarrow \chi\chi$ is in principle available. A measured invisible branching fraction of the Higgs larger than the one expected by SM would be a hint of the presence of dark matter [67, 59].

- **Supersymmetric searches**: Assuming that the WIMP candidate is the *lightest supersymmetric particle* (LSP), the expected signature consists of two or more SM particles produced by the decay of the original SUSY particles and significant missing transverse momentum generated by the LSPs [68, 69].

While supersymmetric searches are based on SUSY predictions to be optimized and interpreted, and the measurement of invisible Higgs branching fraction has to be compared with SM predictions, the other searches use as guidelines simpler models, which do not aim to be complete theories, but minimal extensions of the Standard Model which include DM and are able to guide its discovery. Particular attention is put for these models in the description of the mediator and its couplings to dark matter and standard model particles. In particular, the mediator can be a scalar, a pseudoscalar, a vector, or an axial-vector,

and the couplings are usually chosen in order to give the highest cross sections without violating the results from previous or complementary measurements. More details about the models used for the analysis described in this document, the choice of the parameters, and the phenomenology, are given in Chapter 7. The latest results from the CMS collaboration in the search for DM are shown in the top of Figure 2.12, in the assumption of axial-vector mediator and fixing the coupling of the mediator with quarks g_q to 1, the coupling with DM particles g_{DM} to 0.25, and vetoing coupling to leptons. It can be noted here that di-jet searches have the highest sensitivity, providing the largest exclusion, and being almost independent of the mass of the dark matter particles. Among the mono-X searches, the mono-Jet gives the strongest exclusion limits, followed by mono-γ and mono-Z. It is evident that mono-X searches, which involve the direct production of pairs of dark matter particles, are kinematically limited by the requirement $m_{DM} < \frac{m_{med}}{2}$. These results are also re-interpreted in order to be compared with results from direct detection experiments. In particular, a correspondence between vector and scalar mediator models and spin-independent results from direct detection experiments is given by the following formula:

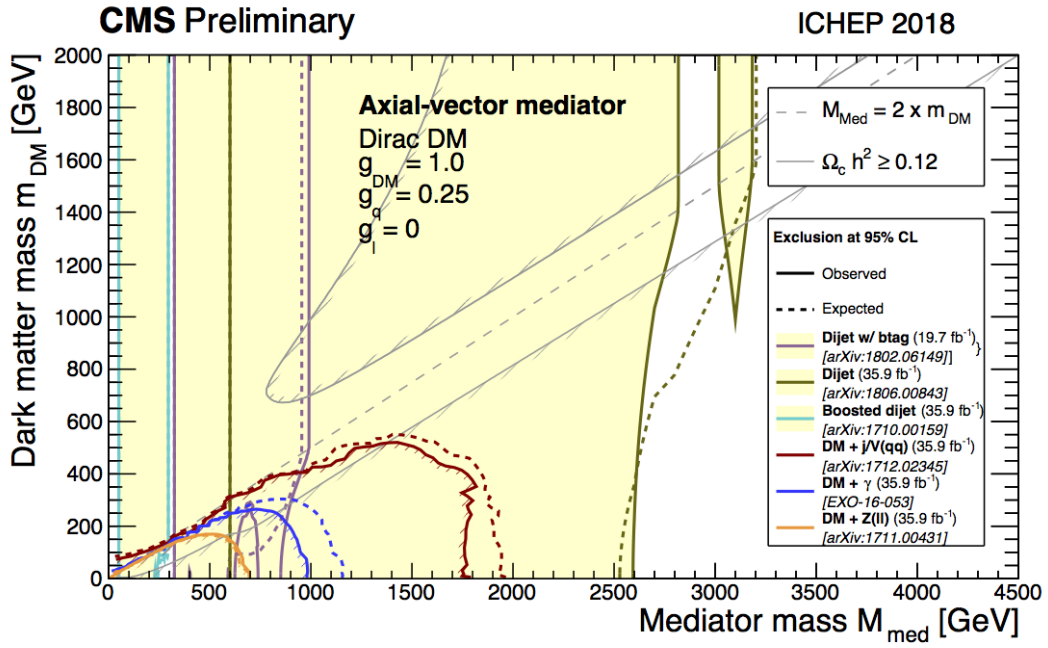
$$\sigma_{SI} = \frac{f^2(g_q) g_{DM}^2 \mu_{n\chi}^2}{\pi M_{med}^4} \text{ [70]} \tag{2.19}$$

where $\mu_{n\chi}^2 = \frac{m_n m_{DM}}{m_n + m_{DM}}$ is DM-nucleon reduced mass with $m_n \sim 0.939$ GeV the nucleon mass. The mediator-nucleon coupling is $f(g_q)$ and depends on the mediator-quark couplings [70]. The correspondence between axial-vector mediator models and spin-dependent results from direct detection experiments is instead given by:

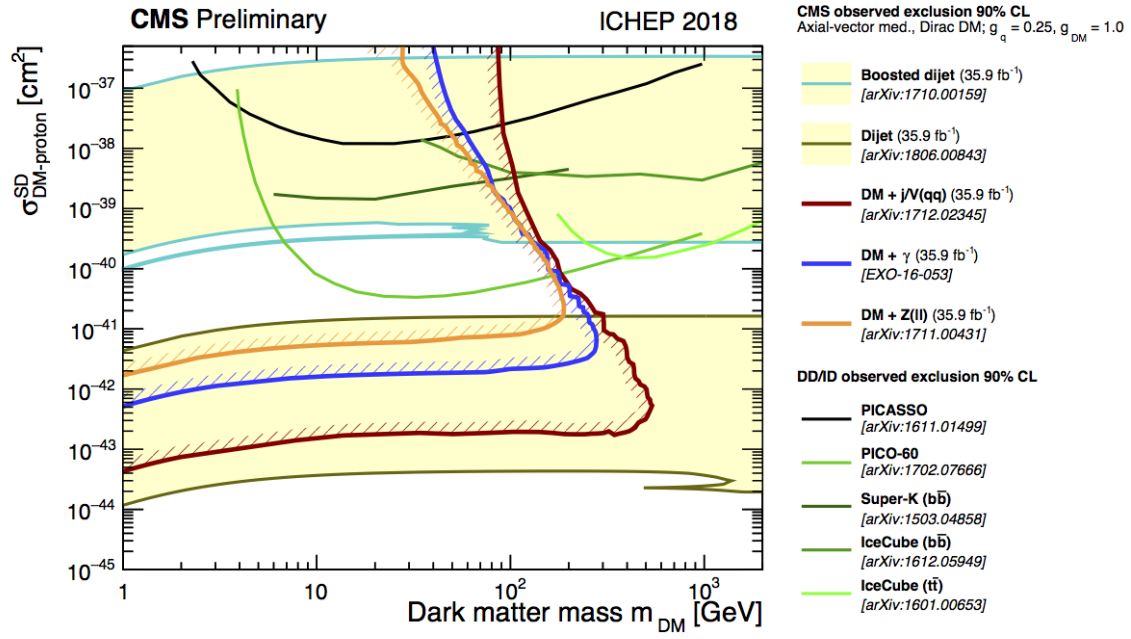
$$\sigma_{SD} = \frac{3 f^2(g_q) g_{DM}^2 \mu_{n\chi}^2}{\pi M_{med}^4} \text{ [70]} \tag{2.20}$$

The comparison between axial-vector mediator simplified models and spin-dependent results from direct detection experiments are shown in the bottom part of Figure 2.12, assuming for collider results the same couplings already mentioned. While direct-detection results are independent of the choice of the coupling, searches at collider need precise assumptions to be re-interpreted as upper limits on the DM-SM interaction cross-section. In particular, the results plotted in Figure 2.12 can drastically change if different couplings are introduced, for example, to allow coupling between the DM mediator and leptons. It is worth noting that collider results are usually given in terms of 95% confidence level exclusion, while direct-detection results in terms of 90% confidence level exclusion. This is taken into account in the re-interpretation of collider searches in terms of SM-DM direct interaction cross section, for which results are shown at 90% confidence level. All the latest results from ATLAS and CMS on the search for DM, with different assumptions on the SM-DM couplings, and the corresponding re-interpretations in terms of direct-detection can be found in the dedicated public pages [71, 72].

The LHCb experiment is designed in order to cover the very forward region, at small angles with respect to the beam pipe. This characteristic makes it a valid instrument to detect the products of collisions on a fixed target, apart from pp collisions. Small amounts of gas, in particular helium, can be injected inside the primary LHC vacuum, allowing proton-Helium collisions, where the protons come from the LHC beams and the Helium acts as a fixed target [73]. p-He collisions with a centre of mass energy of 110 GeV, reproducing the interactions between primary cosmic rays and the interstellar medium, can thus be produced. They are expected to be the main source of cosmic-ray antiprotons in the 10–100 GeV energy range, where precise measurements have been achieved during the



(a) CMS axial-vector results.



(b) CMS spin-dependent interpretation results.

Figure 2.12: Summary plots of the latest results in the search for dark matter from the CMS collaboration (top), and corresponding spin-dependent direct-detection interpretation (bottom). The results are based on simplified models with an axial-vector mediator. The couplings of the mediator are fixed to the following values: coupling with quarks (g_q) is 1, coupling with DM (g_{DM}) is 0.25, and no coupling with leptons is allowed.

last years, as described in Section 2.3.2. The LHCb results for this search show a data-to-simulation yields ratio of 1.19±0.08 for the antiproton production, which if confirmed by

other measurements would indicate an agreement with the indirect searches results, even if at the moment the large systematic uncertainty on the $\bar{p}$ production cross-section in p-He collisions significantly limits the sensitivity of the analysis [73].

Chapter 3

Standard Model

(Non) sono come tu mi vuoi

CCCP, *Sono come tu mi vuoi*

Elementary particles and their interactions are currently best described by the Standard Model (SM) of particle physics. It is a relativistic quantum field theory that coherently accommodates the electromagnetic, the weak, and the strong forces. Over the years it has been tested to an impressive level of accuracy by different experiments and at different energies. The SM is the fundamental tool to predict the expected results of the collisions happening at the LHC and of a particle physics experiment in general, and on the other hand, it is constantly tested by the observations of these same experiments. For this reason, a precise knowledge of the SM is necessary since any discrepancies between its predictions and the results of an experiment are hints of the presence of new physics. In this chapter, an introduction to the basic concepts of the Standard Model is presented. The electroweak symmetry breaking and the Higgs mechanism are also illustrated, as they represent the way in which particles acquire their masses, and given the relevance of the Higgs boson production in this document. To conclude, a review of the main open points of the Standard Model is discussed. A complete and exhaustive description of the Standard Model is beyond the scope of this document but can be found in several books [74].

3.1 Basic Pieces of the Standard Model

The SM is a relativistic quantum field theory based on the group of symmetries $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, which describes three of the four known interactions: electromagnetic, weak, and strong forces. Gravitation, which is much weaker than the other three, cannot be accommodated in the model. The fundamental blocks of this theory are elementary particles and their interactions.

3.1.1 Fundamental Particles of the Standard Model

The fundamental particles of the SM can be categorized into two: on one side the constituents of matter, which are fermions, and on the other the mediators of the interactions, which are gauge bosons. A description of these particles is presented in this paragraph and is illustrated in Figure 3.1.

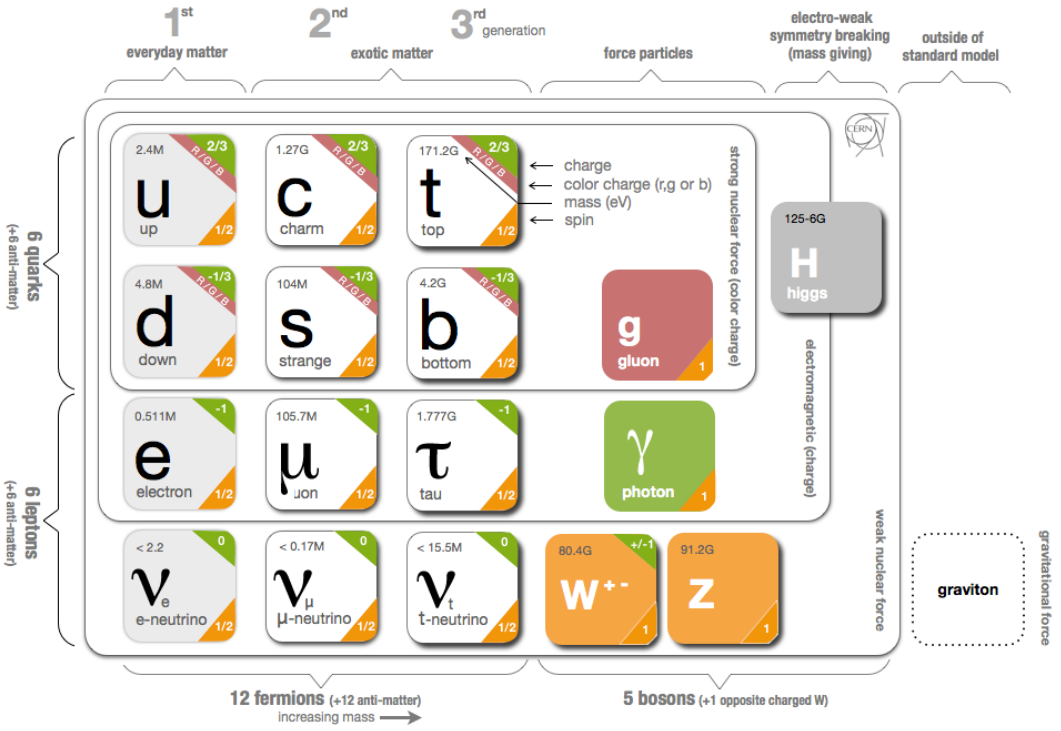


Figure 3.1: The fundamental particles of the Standard Models and their properties [75].

Matter Constituents The constituent of matter are particles of spin $\frac{1}{2}$ called fermions. They are divided into leptons and quarks. While leptons interact only through the electroweak force, quarks are sensitive also to the strong interaction and carry an additional quantum number, called colour. Up to now, quarks have only been observed in bound states, for example as mesons ($q\bar{q}$) or baryons (qqq or $\bar{q}\bar{q}\bar{q}$). Fermions are classified in three families or generations, which differ for their masses. Fermions of the first generation are the lightest ones and constitute ordinary matter, while fermions of the second and third generations present increasing masses and are accessible at higher energies. Inside each family, quarks are further divided into *up* and *down* types. Up quarks have an electromagnetic charge of $+\frac{2}{3}$, while down quarks have an electromagnetic charge of $-\frac{1}{3}$. Leptons are separated in charged leptons, with charge -1, and neutrinos, which are electrically neutral. For each fermion, there exists a corresponding anti-fermion, with the same mass and spin, and opposite electromagnetic charge. It is still not known if neutrinos and anti-neutrinos are the same particles or not.

Interaction Propagators The mediators of the fundamental interactions are particles of spin 1 called gauge bosons. The electromagnetic force is mediated by the photon, which couples to all charged particles. It is massless and electrically neutral. Gluons are the mediators of the strong force and interact with quarks and among themselves. They are massless and electrically neutral and carry the colour charge. Colour has three degrees of freedom (red, blue and green) requiring the presence of eight different gluons. The W^+, W^- and Z bosons are the mediators of the weak force, and interact with quarks, leptons,

and among themselves. The W bosons have electric charge +1 or -1, and are each other's anti-particle, while the Z boson is electrically neutral and is its own anti-particle. They are massive particles and acquire mass through the electroweak spontaneous symmetry breaking mechanism [76, 77].

3.1.2 Interactions in the Standard Model

Interactions in the SM arise by requiring the invariance of its Lagrangian under the so-called gauge symmetry. As an example to illustrate how the introduction of this symmetry naturally induces interaction between particles, the case of the quantum electrodynamics (QED) is presented. The same argument is then extended to the strong interaction, described by the quantum chromodynamics (QCD) and to the electroweak (EWK) interaction.

Lagrangian of the QED

The Dirac Lagrangian describes the motion of a free fermion field $\psi(x)$ in each point of the space-time. It takes the form:

$$\mathcal{L}^{Dirac} = \bar{\psi}(x)(i\gamma^{\mu}\partial_{\mu} - m)\psi(x) \tag{3.1}$$

The idea now is to require this Lagrangian to be invariant under a gauge transformation of the field $\psi(x)$ through a generic function $\alpha(x)$. A gauge transformation is a phase transformation of the field ψ(x), as:

$$\psi(x) \rightarrow \psi'(x) = e^{ie\alpha(x)}\psi(x) \tag{3.2}$$

where e is the dimensionless coupling strength of electrodynamics. It is easy to show that Eq. 3.1 is not invariant under this transformation. Nevertheless, the goal can be achieved by introducing the covariant derivative:

$$\partial_{\mu} \rightarrow \mathcal{D}_{\mu} = \partial_{\mu} + ieA_{\mu} \tag{3.3}$$

where A_{μ} is the gauge field, corresponding to the photon. It is defined by its gauge transformation property:

$$A_{\mu} \rightarrow A'_{\mu} = A_{\mu} - \partial_{\mu}\alpha(x) \tag{3.4}$$

Using now the definition of covariant derivative, Eq. 3.1 takes the form:

$$\mathcal{L}^{Dirac} = \bar{\psi}(x)(i\gamma^{\mu}\partial_{\mu} - m)\psi(x) - e(\bar{\psi}(x)\gamma^{\mu}\psi(x))A_{\mu} \tag{3.5}$$

which is now invariant under gauge transformations. The Lagrangian still cannot describe the free propagation of photons. To do this, a gauge-invariant kinematic term involving A_{μ} has to be introduced. First of all the field strength is defined as:

$$F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu} \tag{3.6}$$

So that the free Lagrangian for the propagation of photons can be described by the following gauge-invariant term:

$$\mathcal{L}^{gauge} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{3.7}$$

By including this last term in Eq. 3.5, one gets the final Lagrangian of the QED:

$$\mathcal{L}^{Dirac} = \bar{\psi}(x)(i\gamma^{\mu}\partial_{\mu} - m)\psi(x) - e(\bar{\psi}(x)\gamma^{\mu}\psi(x))A_{\mu} - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{3.8}$$

Where the first term represents the free propagation of fermions, the second term the interaction of fermions with photons, and the last term the free propagation of the photons. QED thus represents a relatively simple example that shows that starting by the free fermion propagation Lagrangian and requiring it to be invariant under gauge transformations, one gets the interaction between the fermions and the gauge bosons responsible for the mediation of the force. The propagation of the gauge boson is then included by the corresponding gauge-invariant term. QED is described by the U(1) symmetry group, while QCD and the electro-weak unification are more complex interactions and are described by different symmetry groups, but the basic idea is the same.

Lagrangian of the QCD

The strong force is mediated by gluons, and quarks are the only fermions that interact via this force. The same idea used for QED can be applied also to this more complicated case, to obtain the interactions starting from the free Lagrangian. The theory that describes this force is QCD, whose symmetry group is $SU(3)_C$, where C indicates the colour. The generators of this group are eight, corresponding to the eight gluons. Since the generators do not commute to each other ($[T_a,T_b] = i\lambda_{abc}T_c$), QCD is a non-abelian gauge theory. The field that interacts, in this case, is a triplet:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} \tag{3.9}$$

Also in this case the starting point is the Dirac Lagrangian of Eq. 3.1, where now m is a mass matrix, instead of a single number. As already seen, to introduce the interaction in the Lagrangian, one has to require it to be invariant under gauge transformations, which in this case take the form:

$$\psi(x) \rightarrow \psi'(x) = U(\vec{\theta}(x))\psi(x) = (e^{ig\Sigma_a \theta_a(x) T_a})\psi(x) \tag{3.10}$$

In analogy with the QED case, three steps are needed in order to make the Lagrangian invariant under this transformation:

- Introduce the covariant derivative $\mathcal{D}^\mu = \partial^\mu + igT_a\ G_b^\mu\ G_c^\nu$, where g is the dimensionless coupling strength of the strong force and G_b^μ the gauge field
- Define the gauge transformation properties of the G_b^μ gauge field:
  $G_b^\mu \rightarrow G'^\mu_b = G_b^\mu$ - $\partial^\mu\theta_b$ - $g\lambda_{abc}\theta_a G_c^\mu$ (for small θ_a)
- Define the field strength: $F_a^{\mu\nu} = (\partial^\mu G_a^\nu)$ - $(\partial^\nu G_a^\mu)$ - $g\lambda_{abc}G_b^\mu G_c^\nu$

Following these steps, one gets the QCD Lagrangian:

$$\begin{aligned}\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - g(\bar{\psi}\gamma^\mu T_a\psi)G_a^\mu - \frac{1}{4}(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G_{a\nu} - \partial_\nu G_{a\mu}) \\ + \frac{1}{2}g\lambda_{abc}G_b^\mu G_c^\nu(\partial_\mu G_{a\nu} - \partial_\nu G_{a\mu}) - \frac{1}{4}g^2\lambda_{abc}\lambda_{ars}G_b^\mu G_c^\nu G_{r\mu}G_{s\nu}\end{aligned} \tag{3.11}$$

Where:

- $\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi$ describes the free quark propagation
- $g(\bar{\psi}\gamma^\mu T_a\psi)G_a^\mu$ is the quark-gluon interaction term

- $\frac{1}{4}(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G_{a\nu} - \partial_\nu G_{a\mu})$ is the gluon propagation term
- $\frac{1}{2} g \lambda_{abc} G_b^\mu G_c^\nu (\partial_\mu G_{a\nu} - \partial_\nu G_{a\mu})$ is the triple gauge coupling term, which describes the interaction of three gluons
- $\frac{1}{4} g^2 \lambda_{abc} \lambda_{ars} G_b^\mu G_c^\nu G_{r\mu} G_{s\nu}$ is the quadruple gauge coupling term, which describes the interaction of four gluons

It is worth noting the presence of gluons self-coupling terms: such terms were not present in QED for the photons, which are electrically neutral, and this means that gluons carry colour charge. Another fundamental property of QCD is that it is almost a free theory at high energies and is more intense at low energies, or high distances. This is the origin of the so-called colour confinement phenomenon, for which coloured particles cannot be found isolated. A direct consequence of this phenomenon is the *hadronization*, which is the process of formation of hadrons from free quarks or gluons, typically observed at colliders. Here quarks and gluons start to emit other coloured particles, that eventually recombine in colourless hadrons, producing a cascade of particles called hadronic jet.

Lagrangian of the Electroweak Interaction

In the attempt of building a gauge theory able to explain both charged and neutral weak phenomena, like radioactive β decay and neutrino scattering, Glashow [78], Weinberg [79] and Salam [80] independently developed a theory which described the weak interaction and included the QED. In the electroweak theory, the left and right projections of the fields behave differently. They are defined by the chirality operators:

$$\psi_{R,L} = P_{R,L}\psi = \frac{1}{2}(1 \pm \gamma^5)\psi \tag{3.12}$$

The symmetry of the electroweak interaction is $\mathrm{SU(2)_L \otimes U(1)_Y}$. It requires four gauge fields, which are W^1, W^2 and W^3 (for $\mathrm{SU(2)_L}$), and B ($\mathrm{U(1)_Y}$), and four generators, which are the weak isospin T and the weak hypercharge Y. They are related to the electric charge Q by:

$$Y = 2(Q - T^3) \tag{3.13}$$

Fermions associated to the left projection of fields (left-handed fermions) are paired in isospin doublets with $T = 1/2$ and $T^3 = \pm 1/2$, while right-handed fermions are isospin singlets, with $T = 0$ and $T^3 = 0$:

$$\begin{aligned} \psi_L &= \begin{pmatrix} u_L \\ d_L \end{pmatrix} \\ \psi_R &= (u_R), (d_R) \end{aligned} \tag{3.14}$$

As in the previous cases, the interaction enters in the Lagrangian by introducing a gauge transformation and requiring it to be invariant under this transformation, by adding a covariant derivative. The gauge transformation takes the form:

$$\psi \to \psi' = (e^{i\frac{g'}{2}\alpha_0 Y} \cdot e^{ig\alpha_k T^k})\psi \tag{3.15}$$

where the dimensionless coupling strength g and g' are introduced. The covariant derivative is expressed as:

$$\mathcal{D}^\mu = \partial^\mu + igT_k W_k^\mu + i\frac{g'}{2} Y B^\mu \tag{3.16}$$

So that the Lagrangian acquire the following component, related to the electroweak interaction:

$$\mathcal{L}_{EWK}^{interaction} = -\frac{g'}{2}(\bar{\psi}\gamma_\mu Y\psi)B^\mu - g(\bar{\psi}T^k Y\psi)W^{k\mu} \tag{3.17}$$

Here the Lagrangian is expressed as a function of the ψ field and the gauge fields. The gauge fields do not correspond to the fields representing the observable bosons, which are the W^+, W^-, and Z weak bosons, and the photon, represented by the A field (in analogy with the QED). To express the electroweak interaction Lagrangian in terms of the observable bosons, a rotation angle θ_W (the Weinberg angle) between the gauge fields is introduced, such that:

$$\begin{aligned} B &= A\cos\theta_W - Z\sin\theta_W \\ W^3 &= A\sin\theta_W + Z\cos\theta_W \\ W^{1,2} &= \frac{W^- \pm W^+}{\sqrt{2}} \end{aligned} \tag{3.18}$$

The Weinberg angle is defined by the following relation between the coupling strengths:

$$e = g\sin\theta_W = g'\cos\theta_W \tag{3.19}$$

Introducing the notation $\not{X} = \gamma^\mu X_\mu$ and substituting in Eq. 3.17 the relations of Eq. 3.18 and Eq. 3.19, one obtains:

$$\begin{aligned} \mathcal{L}_{EWK}^{interaction} = &-\frac{g}{\sqrt{2}}[\bar{u}_L \not{W}^- d_L + \bar{d}_L \not{W}^+ u_L] - e\bar{\psi}\not{A}Q\psi \\ &-\frac{e}{2\sin\theta_W\cos\theta_W}\bar{\psi}\not{Z}[T^3 - 2Q\sin\theta_W{}^2 - T_L^3\gamma^5]\psi \end{aligned} \tag{3.20}$$

Here, the first term represents the weak charged interaction, with the $W^\pm$ bosons interacting only with the left-handed component of the fields, the second term is the photon-fermion interaction already seen in QED, and the last term describes the weak neutral interaction of the Z boson, which does not distinguish between left-handed and right-handed components. It is worth underlining that the presence of a mass term for gauge bosons is in general prohibited if one wants to keep the Lagrangian invariant under gauge transformations. A mass term like $mA_\mu A^\mu$ would, in fact, violate the gauge invariance, since the A_μ field transforms as defined in Eq. 3.4. In the case of the electroweak interaction, due to the chiral structure of Eq. 3.20, even mass terms for the fermions would violate the gauge invariance. A mechanism is thus required in order to introduce a mass term for the fermions and the weak gauge bosons, but which is able to preserve both the gauge invariance and the chiral symmetry of the theory.

3.1.3 Particles Masses in the Standard Model

The formalism able to introduce a mass term for the weak bosons and the fermions was first introduced by Higgs [76], Englert and Brout [77]. The idea is to introduce a spontaneous symmetry breaking (SSB) of the gauge invariance. This is realized by introducing a new field, denominated Higgs field, and by making the mass terms arise by choosing arbitrarily a minimal configuration of the potential of the field, without putting any terms in the Lagrangian that explicitly violate the gauge invariance. The minimal way to obtain a SSB is to add the following term to the SM Lagrangian:

$$\mathcal{L}_{Higgs} = (D^\mu\phi)^\dagger(D_\mu\phi) - V(\phi) \tag{3.21}$$

where ϕ is a complex scalar field and V(ϕ) is the potential of the ϕ field. The new field must couple to the massive $W^{\pm}$ and Z bosons but not to the massless photon and gluon, so it has to carry no electric or colour charge, and needs at least three degrees of freedom:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{3.22}$$

The potential V(ϕ) takes the form:

$$V(\phi) = \lambda(\phi^\dagger\phi)^2 + \mu^2(\phi^\dagger\phi) \tag{3.23}$$

Here λ has to be positive in order to ensure the presence of a ground state of the potential. As shown in Figure 3.2, a positive value of μ^2 would lead to a single minimum of the potential at $\phi = 0$, while in the case of $\mu^2 < 0$ the potential has a circle of degenerate minima at:

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} \equiv \frac{1}{2}v^2 \tag{3.24}$$

Where v is the vacuum expectation value of the field ϕ:

$$v = \langle 0|\phi|0\rangle = 246 \text{ GeV [81]} \tag{3.25}$$

The spontaneous symmetry breaking can be achieved only in this second case. It consists of choosing a particular ground state around which the Higgs field ϕ is expanded. Typically the selected ground state is:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{3.26}$$

Expanding the Higgs field around this minimum, one gets:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{3.27}$$

By rewriting the Higgs Lagrangian introduced in Eq. 3.21 in terms of the field expansion around the selected ground state, the following expression is obtained:

$$\begin{aligned} \mathcal{L}_{Higgs} = & \left\{ \frac{1}{2}\partial_\mu h \partial^\mu h - \frac{1}{2}2v^2\lambda h^2 \right\} + \left\{ -\frac{1}{3!}6v\lambda h^3 - \frac{1}{4!}6\lambda h^4 \right\} \\ & + \left\{ \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{-\dagger} W^{-\mu} + \frac{1}{2}\frac{v^2 g^2}{4} W_\mu^{+\dagger} W^{+\mu} \right\} \\ & + \left\{ \frac{1}{2}\frac{v^2(g^2+g'^2)}{4} \left(\frac{gW_\mu^3 - g'B^\mu}{\sqrt{g^2+g'^2}} \right)^2 + 0 \left(\frac{g'W_\mu^3 + gB^\mu}{\sqrt{g^2+g'^2}} \right)^2 \right\} + \\ & + \left\{ \frac{1}{4}(2vh + h^2) \left[g^2 W_\mu^- W^{+\mu} + \frac{1}{2}(g^2+g'^2) \left(\frac{W_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}} \right)^2 \right] \right\} \end{aligned} \tag{3.28}$$

Here, the first line comes from the expansion of the potential V(ϕ). There, in the first parentheses, one can find a kinetic term and a mass term, and in the second parentheses two self interacting terms, one corresponding to a triple coupling and one corresponding to a quartic coupling. All these terms are related to the Higgs boson, the particle associated to the excitation of the Higgs field. In particular, the mass of the Higgs boson is given by $m_h = v\sqrt{2\lambda}$ and cannot be predicted by the theory, since λ is a free parameter. The second

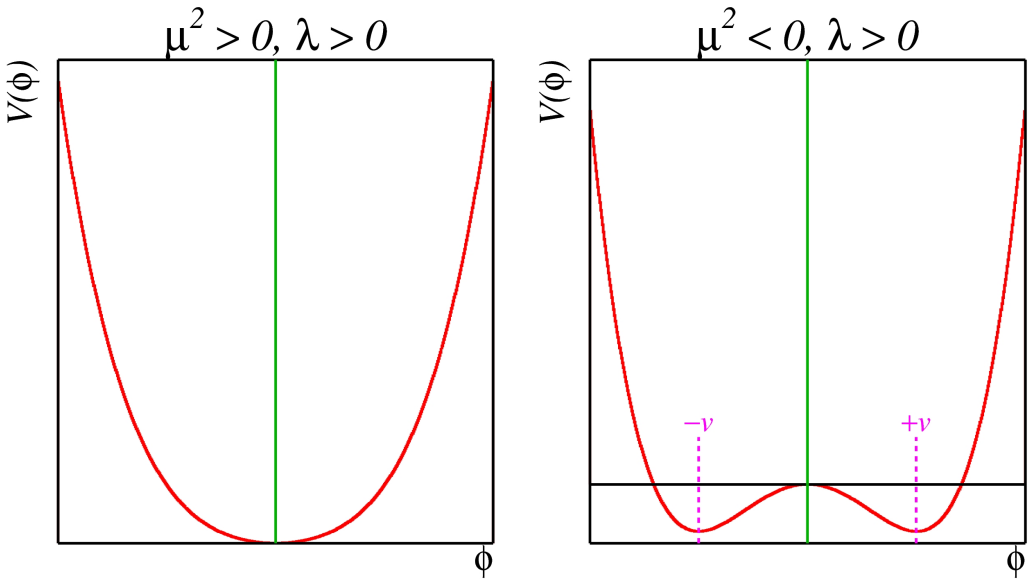


Figure 3.2: Shape of the potential V(ϕ) in the case $\mu^2 > 0$ (left) and in the case $\mu^2 < 0$ (right).

and third lines derive from the kinetic term $(\mathrm{D}^\mu\phi)^\dagger(\mathrm{D}_\mu\phi)$, and contain the mass terms for the gauge bosons. In particular, the mass of the $\mathrm{W}^\pm$ boson is given by:

$$m_\mathrm{W} = \frac{1}{2}vg \tag{3.29}$$

while for the masses of the Z boson and the photon:

$$m_\mathrm{Z} = \frac{1}{2}v\sqrt{g^2+g'^2} \quad \text{where} \quad Z_\mu = \frac{gW^3_\mu - g'B^\mu}{\sqrt{g^2+g'^2}} \tag{3.30}$$

$$m_\mathrm{A} = 0 \quad \text{where} \quad A_\mu = \frac{gW^3_\mu + g'B^\mu}{\sqrt{g^2+g'^2}} \tag{3.31}$$

Remembering Eq. 3.19, the masses of the W and Z boson can be related to each other in the following way:

$$\frac{m_W}{m_Z} = \cos\theta_W \tag{3.32}$$

A very interesting feature of the introduction of the Higgs field is that it is able to produce also the mass terms needed for fermions. For example, for leptons they are obtained by adding the following term to the Lagrangian:

$$\mathcal{L}_{\ell Y} = -G_\ell\left[(\bar{\ell}_L\phi)\ell_R + \bar{\ell}_R(\phi^\dagger\ell_L)\right] \tag{3.33}$$

After the SSB, namely by putting Eq. 3.27 in Eq. 3.33, one gets:

$$\mathcal{L}_{\ell Y} = -\frac{G_\ell}{\sqrt{2}}\left[v(\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L) + (\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L)h\right] = -\frac{G_\ell}{\sqrt{2}}\left[v\bar{\ell}\ell + \bar{\ell}\ell h\right] \tag{3.34}$$

where G_ℓ is known as Yukawa coupling and is chosen in order to generate the correct lepton mass term, so that:

$$m_\ell = \frac{G_{\ell v}}{\sqrt{2}} \tag{3.35}$$

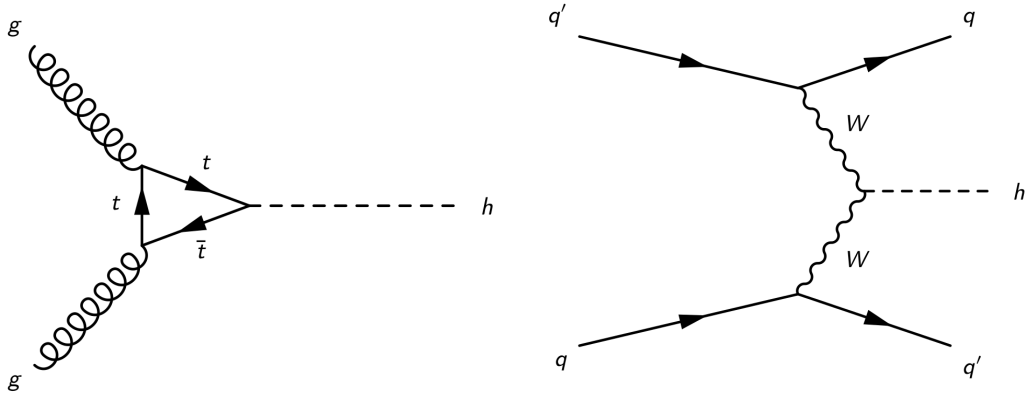
In a similar way, also quarks acquire mass through spontaneous symmetry breaking. It is interesting to underline that a coupling between fermions and the Higgs boson has explicitly been introduced, which is proportional to the mass of the fermions.

3.2 Higgs Boson Discovery

The main target of the first run of data taking of the LHC was the discovery of the Higgs boson. In 2012 the ATLAS [4] and CMS [5] collaborations announced the discovery of a new particle with properties compatible with those predicted by the Standard Model for the Higgs boson, with a mass of:

$$125.09 \pm 0.21 \text{ (stat.)} \pm 0.11 \text{ (syst.) GeV [82].} \tag{3.36}$$

The Higgs production mechanisms at the LHC exploited for this search were the gluon-gluon fusion (with the highest cross section) and the vector boson fusion (easier to tag thanks to the presence of two jets in the final state). A schematic representation of the two processes is presented in Figure 3.4, while in Figure 3.5, the production cross sections for the Higgs boson with different mechanisms are shown, together with the branching fractions predicted by the SM, plotted as a function of the Higgs mass. The Higgs decay

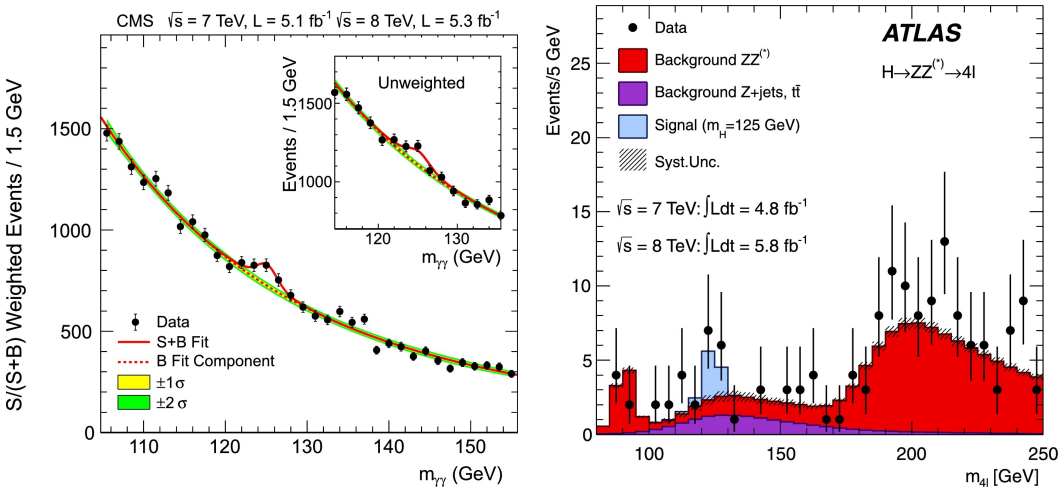


(a) Gluon-gluon fusion.

(b) Vector boson fusion.

Figure 3.3: Feynman diagrams representing the main Higgs production mechanisms at the LHC.

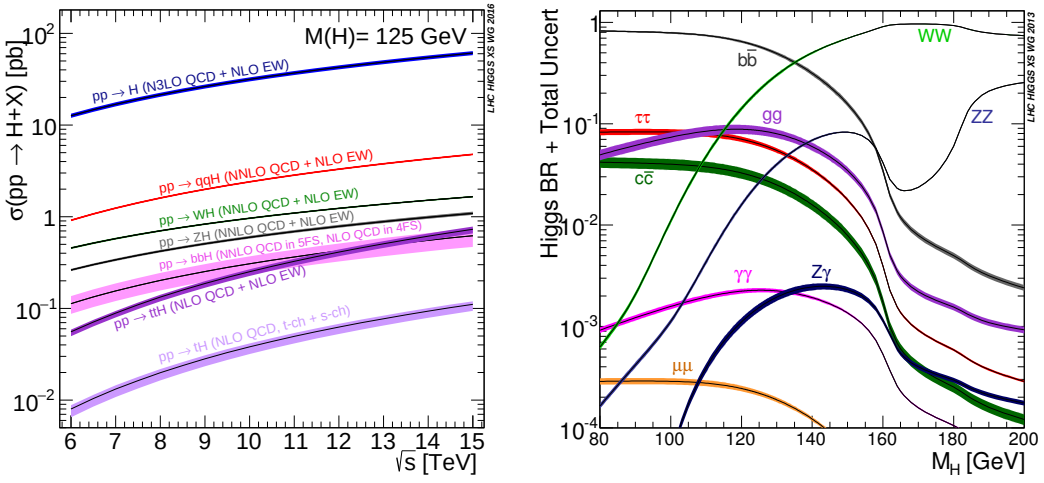
to a pair of photons and to a pair of Z bosons, both decaying to a pair of leptons have been used for the discovery. The $h \to \gamma\gamma$ final state ensures a good invariant mass resolution, allowing to distinguish the peak of events produced by the Higgs boson from the large amount of background. The $h \to ZZ^* \to 4\ell$ decay, on top of a precise invariant mass reconstruction, takes profit from the low number of processes producing the same final state to give a very clean final state. During the Run 2 of the LHC, exploiting the higher amount of data collected and the higher production cross section at 13 TeV with respect to 8 TeV, more decay channels of the Higgs boson have been observed, including the



(a) h→ γγ. (b) h→ZZ* →4ℓ.

Figure 3.4: Invariant mass distributions in the h→ γγ search at CMS (left) and in the h→ZZ* →4ℓ search at ATLAS (right). In both cases, an excess of events with respect to the expectations in the background-only hypothesis is observed at about 125 GeV.

h → WW* [83, 84], h → ττ [85], and h → bb [86]. This allows also to observe other production mechanisms, including production associated with a vector boson or with a pair of top quarks [87, 88]. In particular with a cross-section of about 0.5 pb [89], the ttH associated production represents the current limit of cross-sections that can be observed at the LHC.



(a) Higgs production cross section. (b) Higgs branching fractions.

Figure 3.5: Production cross sections at the LHC for a Higgs boson of mass 125 GeV as a function of the centre of mass energy (left) and Higgs boson branching fractions as a function of its mass (right).

3.3 Open Points of the Standard Model

The discovery of the Higgs boson by the ATLAS [4] and CMS [5] collaborations in 2012 is just the last of a series of impressive achievements the Standard Model showed to be capable of. Nevertheless, it cannot be considered as the ultimate theory of particle physics, but an effective model valid at energies up to the electroweak scale, where different profound theoretical ideas are put together in order to reproduce the experimental data. As already seen in this chapter, the Standard Model is in fact based on:

- Quantum Field theory to provide a fundamental description of particles and their interactions
- The Dirac equation of relativistic quantum mechanics to describe the dynamics of fermions
- Local gauge invariance to get the actual nature of the interactions
- The Higgs mechanism of SSB to give mass to fermions and gauge bosons
- A large set of experimental results that guided the way in which the former ingredients had to be combined together to build the model

This leaves many unanswered questions; a short list of the ones that are currently the subject of experimental research is presented in this section [90].

What is Dark Matter? Since the 1930s it has been known that an important part of the mass of the Universe does not consist of stars, as previously thought. A long series of compelling evidence of the existence of dark matter started to grow since then, as already presented in more detail in Chapter 2. The present results suggest that only 5% of the mass-energy budget of the Universe is made of ordinary matter, while about 25% consists of dark matter, and the remaining is dark energy [8]. The hypothesis that dark matter is composed by particles described by the Standard Model is excluded, so in the assumption that dark matter is made of particles, an extension of the SM in order to account for the constituents of dark matter is needed.

Is the Higgs boson the only responsible for Electroweak Spontaneous Symmetry Breaking? The 125 GeV Higgs boson couplings with elementary particles show it is undoubtedly involved in the process of Electroweak Spontaneous Symmetry Breaking (EWSB) and in giving mass to other particles. Nevertheless, it is still to be verified whether it is an elementary particle or not and if there are additional Higgs bosons taking part in the process.

Hierarchy Problem and Fine Tuning The Standard Model is a renormalizable theory, which means that its predictions are expressed as expansion series. In particular, higher order corrections to the Higgs boson mass due to the interaction with fermions can become very large at very high energies, close to Planck scale ($\Lambda_p \sim 10^{19}$ GeV). This issue goes under the name of Hierarchy Problem. If no physics beyond the SM exists, in order to keep the mass of the Higgs boson finite and of the order of the electroweak scale, fine-tuning of these contributions is required, such that they cancel to a high degree of precision. One of the main motivations for physics beyond the SM is to avoid this fine-tuning. Models that do not need this tuning, introduce new particles with masses of the order of the TeV and which couple to the Higgs boson, so that they can be explored at the LHC.

Why there is an asymmetry between matter and antimatter? Despite matter and antimatter have almost specular properties, the former is much more abundant than the latter in Universe. To explain this observation, the introduction of new physics to provide baryon number violation, CP violation, and out-of-equilibrium physics in the early Universe is required.

What is the origin of particle mass hierarchies and mixing angles? The masses of quarks, leptons, and neutrinos are not predicted by the SM, and their pattern is still a mystery. Also the mixing angles are introduced empirically, even if they represent fundamental parameters of particle physics. Models that introduce new particles with flavour-dependent couplings have been proposed as extensions of the SM able to explain those observations.

Gravity The gravitation force is responsible for the formation of the macroscopic structure of the Universe. Despite this, it is much less intense than the other three fundamental forces currently known and is not included in the Standard Model. Even if its contribution at the electroweak energy scale is basically negligible, its role is expected to be much more relevant at the Planck scale, so that a more complete theory able to coherently describe all the four fundamental interactions will be needed.

Chapter 4

The LHC and the CMS Experiment

Tecnica d'acciaio
Scienza armata cemento

CSI, *Unità di produzione*

The data used in this document have been produced through proton-proton collisions with a centre of mass energy of 13 TeV by the Large Hadron Collider (LHC) during 2016. The products of such collisions have been detected by the Compact Muon Solenoid (CMS), for a total integrated luminosity of 35.9 fb^{-1}. In this chapter, a description of the properties, the performances, and the phenomenology of the LHC is presented, followed by an overview of the CMS experiment, its sub-detectors, and the introduction of some physical quantities of particular interest. More complete discussions on the LHC accelerator [91, 92] and on the CMS experiment [93, 94] can be found in the corresponding technical design reports.

4.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a superconducting proton and heavy ions accelerator and collider, located in the underground tunnel originally built for the Large Electron Positron collider (LEP) [95]. The main purpose of the LHC physics program is the investigation of the electroweak spontaneous symmetry breaking mechanism. An important part of this goal has been achieved in 2012, with the discovery of a new particle with characteristics compatible with the Higgs boson [4, 5]. Additionally, the LHC can give an insight into other fundamental physics matters. First of all, it allows testing the Standard Model at the energy scale of the order of the TeV through precise measurements of QCD, EWK and flavour physics processes. These measurements give sensitivity also to possible new physics, which is another main topic of the LHC program. In particular, a great effort is dedicated to the search of supersymmetric particles with masses of the order of the TeV and to the search of dark matter. Finally, the physics program of the LHC includes the study of ions collisions, that can take advantage of the unprecedented centre of mass energy, more than one order of magnitude larger than the 200 GeV attainable at the Relativistic Heavy Ion Collider (RHIC) [96].

4.1.1 LHC Properties

The LHC is a 27 km long circular accelerator. To bend the protons or the heavy ions, it uses 1232 superconducting Niobium-Titanium dipole magnets able to generate a magnetic field of 8.3 T and operating at a temperature of 1.9 K. Since it collides protons against protons (or heavy ions of the same charge), it requires two separate vacuum tubes with opposite magnetic dipole fields, in order to bend the two beams in opposite directions. The two beams share the same beam pipe only in the interaction regions. The proton beams are accelerated in different stages before being injected into the LHC. The full acceleration chain is shown in Figure 4.1 and includes the Super Proton Synchrotron (SPS), which prepares the beams for the injection in the LHC, rising their energy up to 450 GeV. Once in the LHC, the acceleration of the beams up to the nominal energy is performed by 16 superconducting radiofrequency cavities. The usage of this technology makes the protons inside each beam grouping in bunches. The nominal number of bunches is 2808, with a bunch spacing of 25 ns, corresponding to a bunch crossing rate of 40 MHz. There are four interaction points, corresponding to the four big experiments located along the ring: ATLAS, CMS, LHCb and ALICE. ATLAS and CMS are two multipurpose experiments, built with complementary technologies and designs. A description of CMS will be presented in the next section of this chapter. LHCb has been specifically designed to perform precise measurements of CP violation, in order to detect possible evidence of new physics [97]. ALICE is dedicated to the study of heavy ions collisions, where the extreme energy densities lead to the formation of a new state of matter, known as quark-gluon plasma [98].

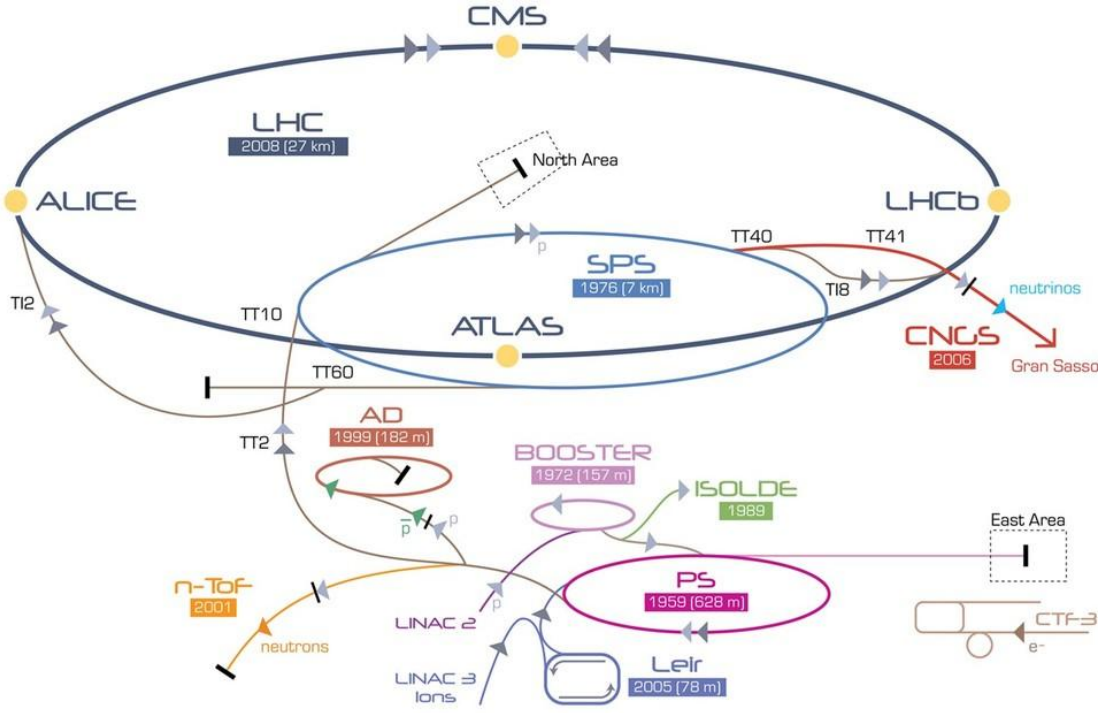


Figure 4.1: Schematic view of the CERN accelerator complex.

4.1.2 LHC Performances

The instantaneous luminosity and the collisions centre of mass energy make the LHC an ideal machine to perform precision measurements, study rare processes and look for new

physics involving heavy particles. The instantaneous luminosity indicates the frequency of proton-proton interactions, such that a higher luminosity reflects in a larger number of collisions. In particular, given a process with a cross-section σ, the rate of events production for this process is given by:

$$R = \mathcal{L} \times \sigma \ [events/s] \tag{4.1}$$

The instantaneous luminosity is defined by several properties of the collider:

$$\mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi\varepsilon_n \beta^*} F \tag{4.2}$$

where:

- γ is the Lorentz factor
- f is the revolution frequency
- k_B is the number of proton bunches per beam
- N_P is the number of protons per bunch
- ε_n is the normalized transverse emittance
- β^* is the betatron function at the interaction point
- F is the reduction factor due to the beam crossing angle at the interaction point

The LHC has already exceeded by a factor 2 the design instantaneous luminosity of $\mathcal{L} = 10^{34}$ cm^{-2} s^{-1} $= 10^{-2}$ pb s^{-1}. One can also define the integrated luminosity:

$$L = \int \mathcal{L} dt \tag{4.3}$$

Where the integration is on the total time of data taking. Considering the integrated luminosity, it is possible to have an estimation of the number of events of a process produced in the whole data taking period:

$$N = L \times \sigma \ [events] \tag{4.4}$$

This means that in 2016, with a total integrated luminosity of $L \sim 40$ fb^{-1}(see the orange line in Figure 4.2), about 40 events of a process with a cross-section of 1 fb have been produced. To achieve such a high luminosity, the LHC beams are set in order to produce more than one proton-proton interaction at each bunch crossing. Typically just one of these collisions generates a hard scattering interesting for physics, while the others simply overlap in the same event. For this reason, it is common to refer to these secondary interaction vertices as *pile-up*. During 2016 the average number of interactions per bunch crossing was 27, as shown in Figure 4.3. An example of an event with a large amount of pile-up is illustrated in Figure 4.4, with the individual vertices of interaction highlighted. The other interesting feature of the LHC is the unprecedented centre of mass energy reached. The design value is of $\sqrt{s} = 14$ TeV and will be reached during the data taking period starting in 2020. In the first part of the first data taking run (Run 1), started in 2010, the centre of mass energy was set to 7 TeV and then increased to 8 TeV. After the long shutdown started in 2013, in 2015 the LHC began its second data taking run (Run 2) at a centre of mass of 13 TeV.

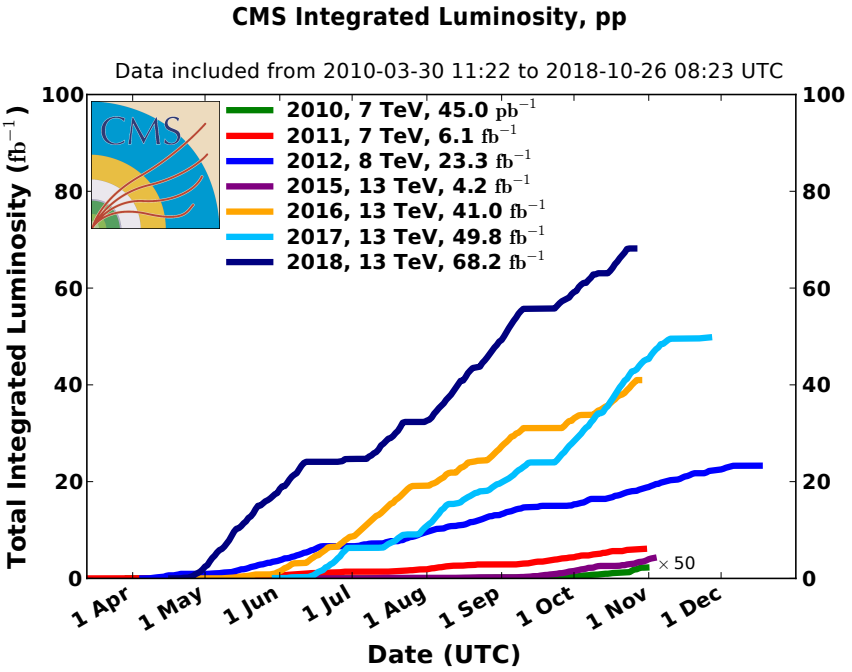


Figure 4.2: Summary of the integrated luminosity collected by CMS, year per year.

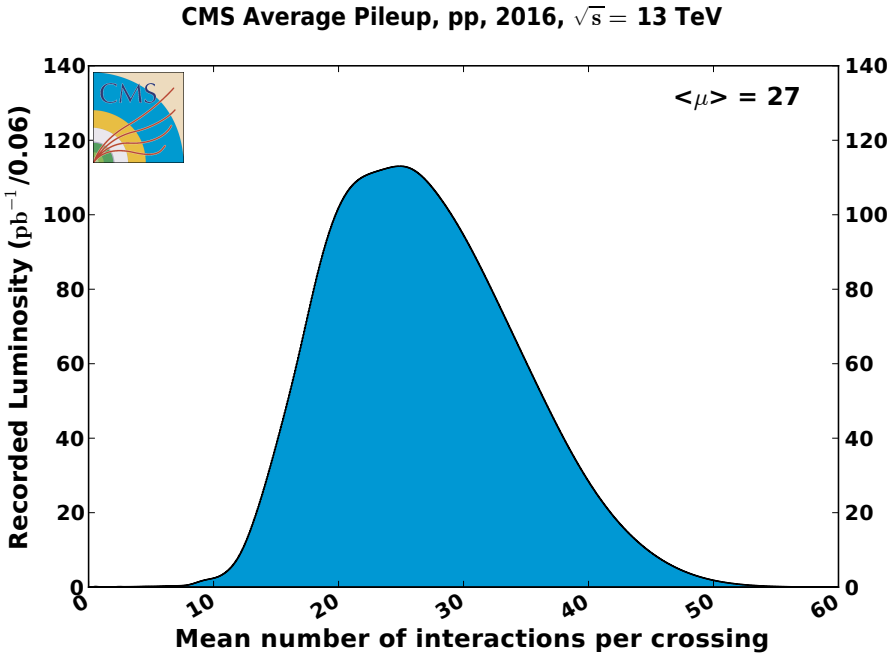


Figure 4.3: Distribution of pile-up during 2016 data taking. The average number of interactions per bunch crossing was 27.

4.1.3 LHC Phenomenology

Protons are not elementary particles, so they are not taking part in the collisions at the LHC as a whole. In fact, the interacting particles are the partons that constitute the hadron. A proton is composed of three valence quarks (2 quarks *up* and one quark *down*).

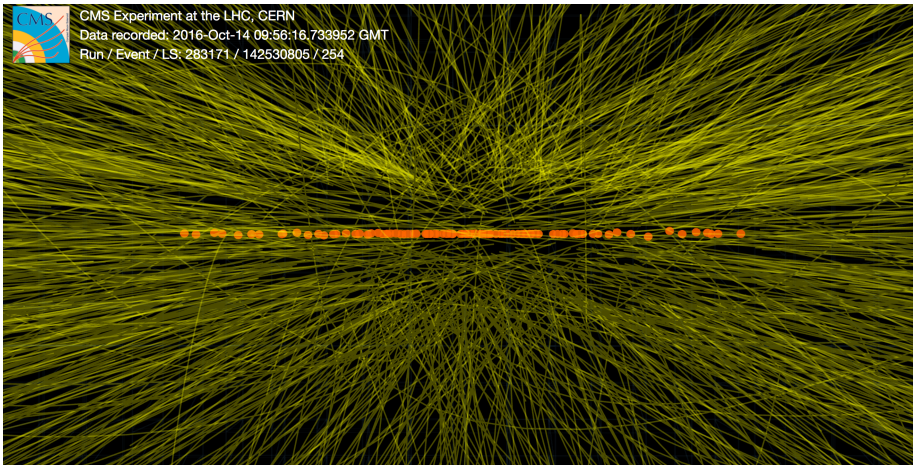


Figure 4.4: Event display of a proton-proton collision taken during the high luminosity fill of 2016, when the average pile-up was around 100 [99].

These quarks interact among themselves by the exchange of gluons, which can self-interact producing more gluons, or producing additional quark-antiquark pairs, which form the so-called *sea*. The momentum of the hadron is shared among the valence quarks, the gluons, and the sea. The fraction of the hadron momentum carried by each parton is not fixed, instead it is described by the Parton Distribution Functions (PDFs), which determine the probability of finding a parton of a given momentum $p_i = xP$, where x is the fraction of the total hadron momentum P carried by the parton, as schematically illustrated in Figure 4.5. At low energies, the valence quarks carry almost all the momentum, but in

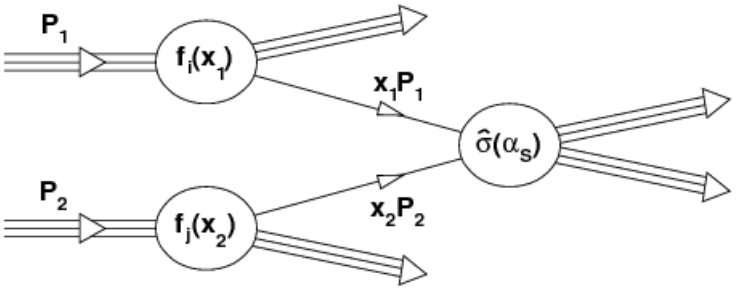


Figure 4.5: Schematic representation of the interaction between two partons in a proton-proton collision [100].

collisions with large energy transfer ($Q^2 \geq 1$ GeV), the gluons and the quarks in the *sea* start to become accessible. The PDFs depend on the energy transfer Q^2 since partons with high x tend to radiate, producing new partons with low x and losing energy themselves, getting lower x values. Once the PDFs are measured at a selected Q^2, their behaviour at different energies can be extrapolated by using QCD, as shown in Figure 4.6. The hard interaction in proton-proton collisions happens only between one parton from each proton, but in any case several other interactions among the remnants of the hadrons may occur. Since typically they are much softer than the main one, they do not produce easily

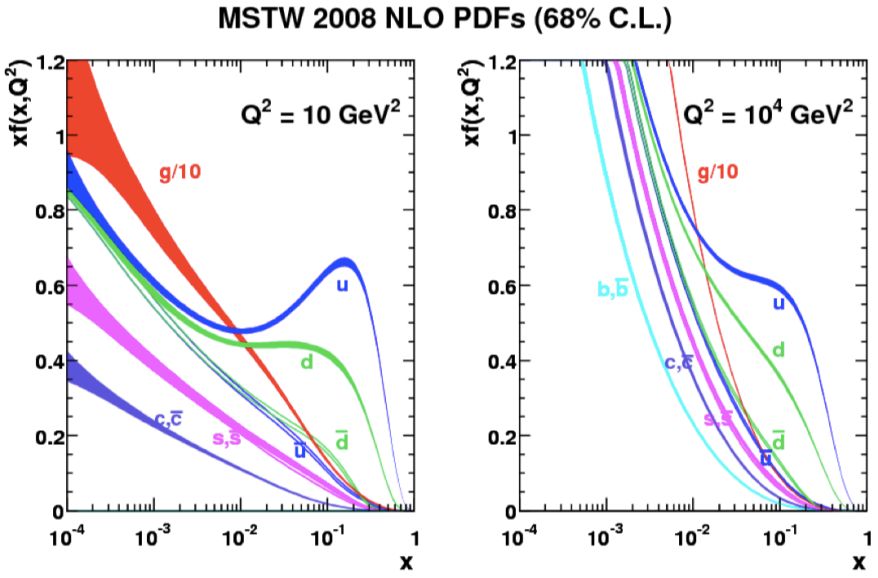


Figure 4.6: Parton distribution functions values from the MSTW group [101]. The two plots show PDFs at two different energy scales: $Q^2 = 10$ GeV^2 on the left and $Q^2 = 10^4$ GeV^2 on the right.

identifiable additional particles in the event, but can contribute to the total amount of scattered energy. Such interactions are denominated *underlying event*.

4.1.4 LHC Kinematics

The existence of a substructure in protons and the PDFs imply that a collision between two protons with the same kinetic energy may results in an interaction between two partons with different momenta. For this reason the center of mass of the collision may not be at rest in the laboratory reference system, but may acquire a boost along the beam axis (longitudinal boost). Kinematic variables which are invariant under this kind of transformation are conveniently used at hadron colliders. In particular, defining the z-axis as the longitudinal axis, the y-axis as the vertical axis, and the x-axis as the axis pointing to the center of the LHC, it is easy to show that the *transverse momentum* $\mathrm{p_T} = \sqrt{p_x^2 + p_y^2}$ of a particle is invariant under longitudinal boosts. To describe the angle between the direction of a particle and the beam axis, it is common to define the rapidity:

$$Y = \frac{1}{2}\, ln\left(\frac{E + p_z}{E - p_z}\right) \tag{4.5}$$

The rapidity transforms under longitudinal boosts as: $Y' = Y - \gamma$, where γ is the longitudinal boost, such that differences in rapidity do not depend on such transformation. For massless particles, or for particles with momenta significantly larger than their rest mass, the rapidity is equivalent to the pseudo-rapidity:

$$\eta = -\log\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{4.6}$$

where θ is the angle between the particle direction and the beam axis. Also in this case it is true that differences in η are invariant under longitudinal boosts.

4.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a general purpose detector installed at the interaction point 5 along the LHC tunnel. It is designed to cover almost hermetically the area around the collision point, with a symmetrically cylindric barrel closed by two endcaps. A schematic representation of CMS is illustrated in Figure 4.7. The characteristics and

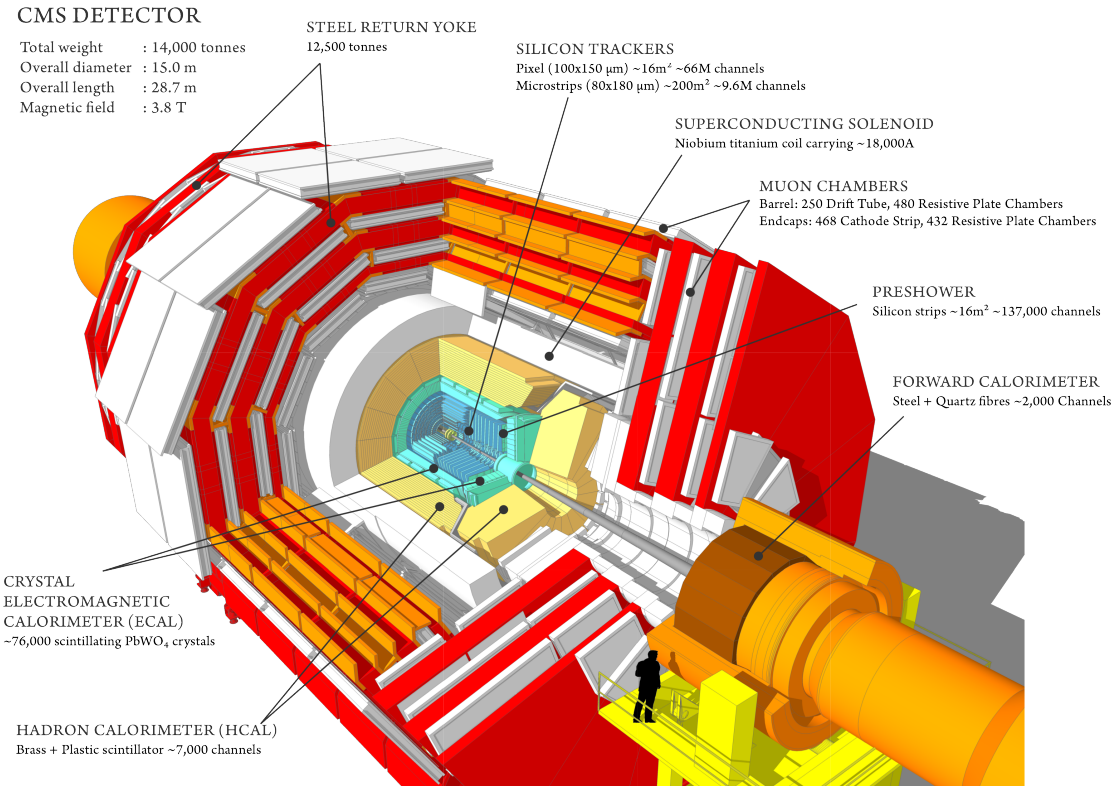


Figure 4.7: Schematic view of the CMS experiment, showing the main components and sub-detectors systems.

the performances that CMS needs to meet the goals of the LHC Physics program can be summarized as follows [94]:

- Good muon identification and momentum resolution, 1% di-muon mass resolution at 100 GeV/c^2 and the ability to determine unambiguously the charge of muons with momentum up to 1 TeV/c.
- Good charged particle momentum resolution and reconstruction efficiency: efficient triggering and offline tagging of τ leptons and b-jets, and high vertex reconstruction efficiency.
- Good electromagnetic energy resolution, di-photon and di-electron mass resolution, hermeticity and efficient photon and lepton isolation at high luminosities.
- Good E_T^{miss} and di-jet mass resolution requiring hadron calorimeters with large hermetic coverage and fine lateral segmentation.

The central feature of CMS, required to achieve the target muon momentum resolution, is a superconducting solenoid, with a diameter of 5.9 m and a length of 13 m, which produces

a magnetic field of 3.8 T. A silicon tracker, an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL) are located inside the solenoid, making CMS a relatively compact detector. Outside the solenoid, the return magnetic field is large enough to saturate the 1.5 m of iron of the holding structure that contains the outer muon tracking detectors. The compactness of the CMS detector is obtained thanks to the high density of the materials employed, so that its weight is about 12500 t, for a length of 22 m and a diameter of 15 m. The coordinate system adopted by CMS has the origin centred at the nominal collision point. The y-axis points vertically upwards, while the x-axis points radially inward towards the LHC ring centre and the z-axis points along the beam direction, towards the Jura mountains from LHC Point 5. A representation of the CMS coordinate system is shown in Figure 4.8. The main four CMS sub-detectors, namely (from the inner-

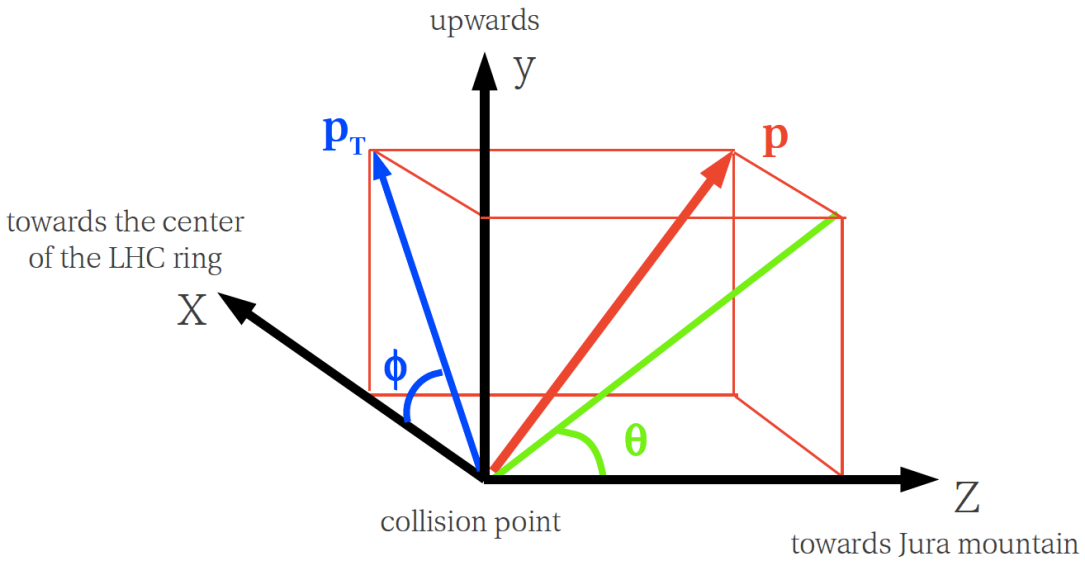


Figure 4.8: Representation of the CMS coordinate system.

most to the outermost) the silicon tracker, the ECAL, the HCAL, and the muon system will be described in the following sections of this chapter, together with the trigger system, necessary to select the interesting events.

4.2.1 The Tracker

The purpose of the CMS tracking system is to reconstruct the primary and secondary interaction vertices and to reconstruct the charged tracks and measure their momenta. It is composed of silicon devices, it is 5.4 m long and has an external diameter of 2.4 m. The design is optimized to reduce the detector occupancy, namely the fraction of active channels per LHC bunch crossing, and to have a small material budget without affecting the performances. The layout of the CMS tracker is shown in Figure 4.9. Closest to the beamline, the Silicon Pixel Detector (SPD) is used to reconstruct the secondary vertices of the interactions. It is made of three layers of silicon pixel semiconductors, located at 4.4 cm, 7.2 cm, and 10.2 cm from the beam line in the barrel, and two disks of 6 cm and 15 cm radius, located at $|z| = 34.5$ cm and $|z| = 46.5$ cm, in each endcap, as shown in Figure 4.10. Each pixel has a surface of 100×150 μm^2 and a thickness of 250 μm. The SPD covers the region $|\eta| < 2.6$, with an area of about 1 m^2 and a total of 66 million pixels. The SPD is surrounded by the Silicon Strip Detector (SSD), made of silicon strips with a pitch which varies between 80 μm and 180 μm. 10 cylindrical layers in the barrel and 9 disks

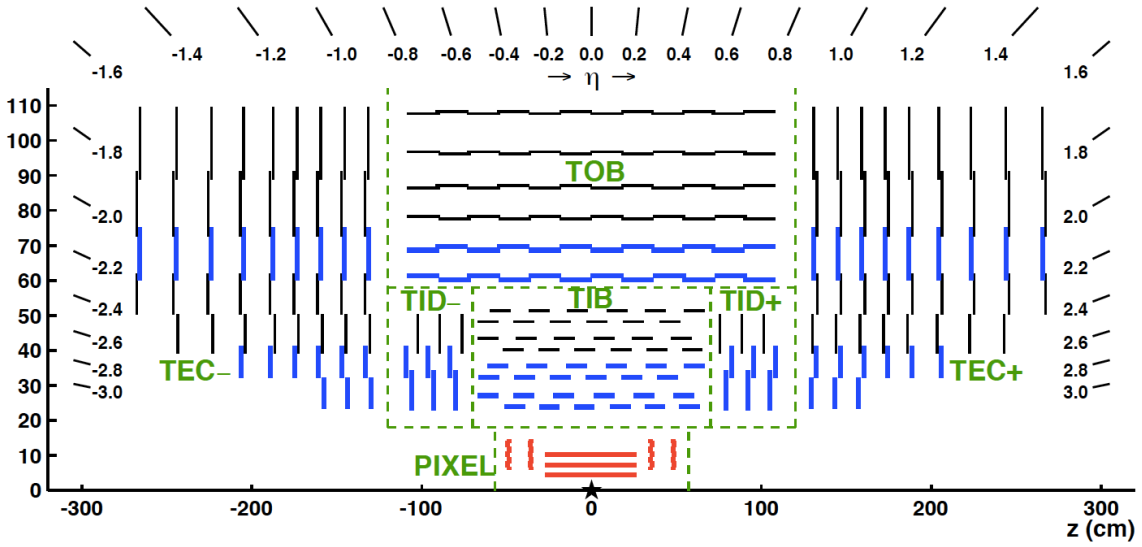


Figure 4.9: View of the CMS tracker in the rz-plane. Each line in the strip tracker represents a silicon strip detector, whereas lines in the pixel detector represent ladders and petals on which the detectors are mounted in the barrel and endcaps, respectively.

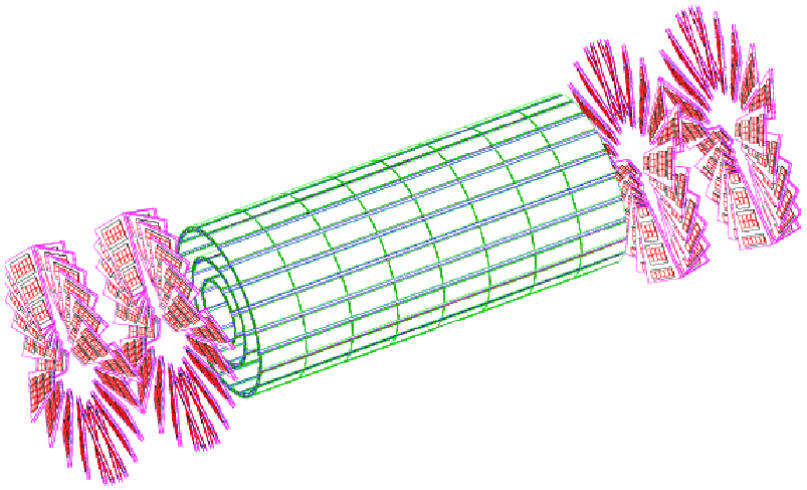


Figure 4.10: Illustration of the CMS pixel detector. The three barrel layers are organized in ladders, the four endcap disks in petals [93].

in each endcap allow covering the region in $|\eta| < 2.5$. The barrel is divided into a Tracker Inner Barrel (TIB) and a Tracker Outer Barrel (TOB). To avoid excessively shallow track crossing angles, the TIB is shorter than the TOB, to accommodate 3 Tracker Inner Disks (TID) at each side of the TIB. The SSD has a total of 9.6 million silicon strips, covering an area of about 200 m^2. The spatial resolution of the CMS Tracker is presented in Table 4.1, separately for the SPD and the SSD. This allows reaching a momentum resolution, for $|\eta| < 1.6$ of:

$$\frac{\sigma(p_T)}{p_T} = \left(\frac{p_T}{GeV/c}\right) \cdot 0.015\% \oplus 0.5\% \tag{4.7}$$

Detector	Resolution (r,ϕ)	Resolution (z)
SPD (barrel)	15 μm	11-17 μm
SPD (endcap)	15 μm	90 μm
SSD	15 μm	1 mm

Table 4.1: Space resolution for the CMS Tracker sub-detectors, in the (r,ϕ) plane and in the z-direction.

The resolution degrades when $|\eta|$ increases, reaching, at $|\eta| = 2.5$ the value:

$$\frac{\sigma(p_T)}{p_T} = \left(\frac{p_T}{GeV/c}\right) \cdot 0.060\% \oplus 0.5\% \tag{4.8}$$

The material budget, shown in Figure 4.11, increases from 0.4 X_0 at $\eta = 0$, reaches its maximum in the transition between endcap and barrel at $|\eta| = 1.6$, with 1.8 X_0, and then decreases again, and represents the main source of uncertainty in calorimetric measurements of electrons and photons (which convert into e^+e^- pairs).

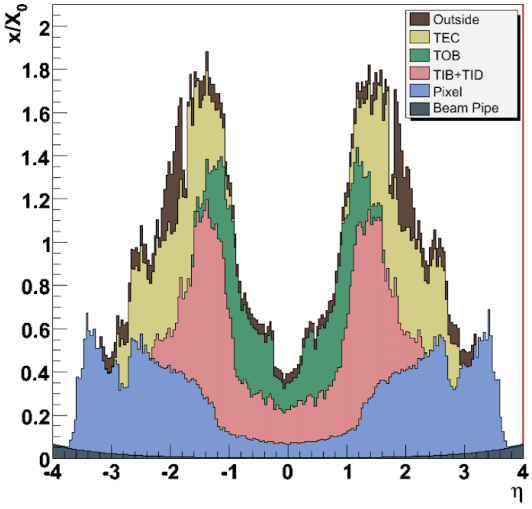
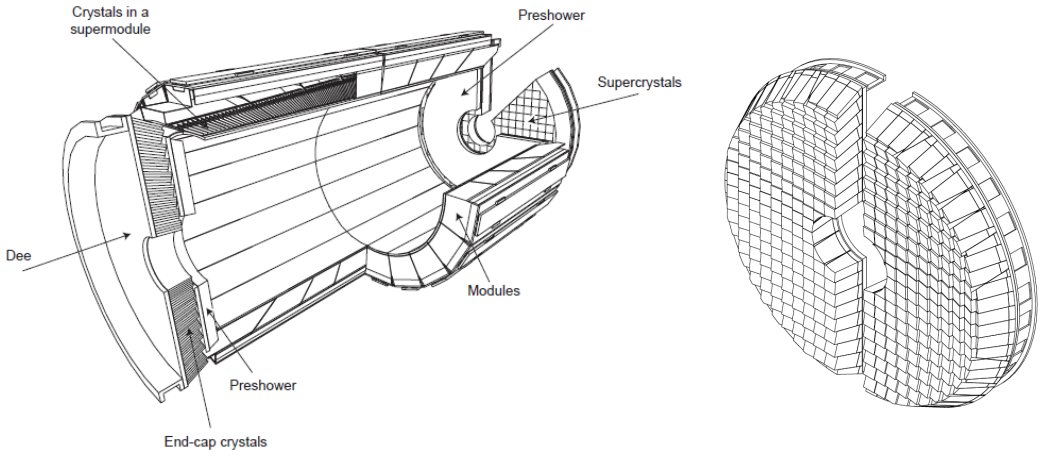


Figure 4.11: CMS Tracker material budget, expressed in radiation length X_0, as a function of η. The maximum is reached in the transition region between the barrel and the endcaps.

4.2.2 The Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is used to precisely measure the energy of electrons and photons, and to measure the fraction of the energy of a jet coming from the production of an electromagnetic shower. It is a hermetic and homogeneous calorimeter composed of lead tungstate ($PbW0_4$) scintillating crystals. The barrel, covering the area $|\eta| < 1.479$, includes 61200 crystals located at a radius of 1.29 m with respect to be beam pipe, while each of the two endcaps covers the region $1.479 < |\eta| < 3$ and is made of 7324 crystals, located at a distance of 3.14 m from the interaction vertex along z. The structure of the ECAL is shown in Figure 4.12. All the crystals in the ECAL are identical, with a front cross-section of 22×22 mm^2 and a length of 230 mm, corresponding to 25.8 radiation lengths. The crystals are truncated pyramid-shaped and mounted in a geometry which



(a) ECAL. (b) ECAL Endcap.

Figure 4.12: Illustration of the CMS ECAL geometry. On the left (a) the full ECAL layout is presented, on the right (b) the structure of an endcap is highlighted.

is off-pointing with respect to the mean position of the primary interaction vertex, with a tilt of 3° both in ϕ and η, as shown in Figure 4.13. This avoids the possibility of a particle to pass between two adjacent crystals. In the barrel, the granularity is 360-fold

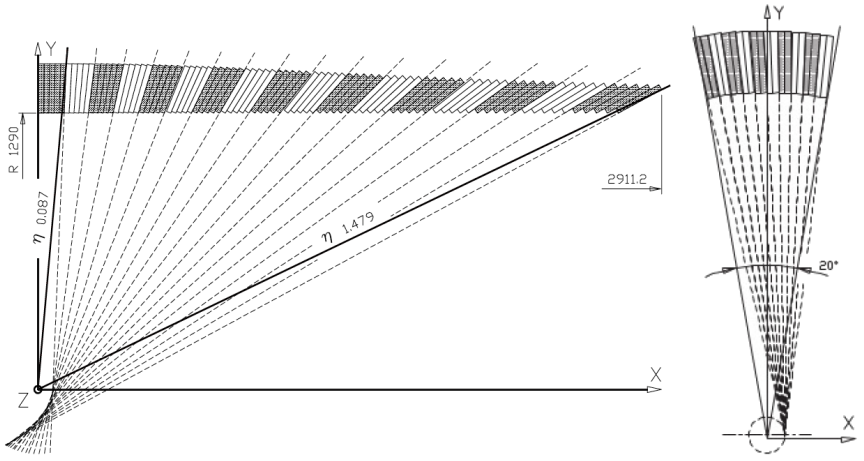


Figure 4.13: Schematic representation of the CMS ECAL crystals disposal. The tilt in η is shown on the left, the tilt in ϕ on the right.

in ϕ, such that every crystal corresponds to 1°, and (2×85)-fold in η. The choice of using lead tungstate for the ECAL crystals was dictated by several properties of the material: it is dense (8.2 g/cm^2), has short radiation length (0.89 cm) and Molière radius (21.9 mm), is fast responding (80% of the light is emitted within 25 ns) and is radiation hard. On the other hand, it provides a low light yields (50 photons per MeV), so that photodetectors with high intrinsic gain and able to operate in a high magnetic field are required. Silicon avalanche photodiodes (APDs) are used in the barrel and vacuum phototriodes (VPTs) in the endcaps. To reduce the probability of very boosted pions decaying to pairs of photons ($\pi^0 \to \gamma\gamma$) to be mis-identified as individual photons, a preshower is located in front of the ECAL endcaps. It is a sampling device, made of two layers of silicon strips with a pitch of 1.9 mm (so more granular than the ECAL itself), the first one located behind a disk of

lead absorber of 2 X_0 thickness, the second one behind a lead disk of 3 X_0 thickness. The lead initiates the showering process for the photons, which can be individually identified thanks to the fine granularity of the silicon detectors. The energy resolution of the ECAL is given by three terms:

$$\frac{\sigma(\mathrm{E})}{\mathrm{E}} = \frac{1}{\sqrt{\mathrm{E/GeV}}} \cdot 2.8\% \oplus \frac{1}{\mathrm{E/GeV}} \cdot 12\% \oplus 0.3\% \tag{4.9}$$

The first term takes into account the stochastic nature of the scintillation and showering processes, the second one is due to the electronic noise, and the third one is related to detector non-uniformity and calibration uncertainty.

4.2.3 The Hadronic Calorimeter

The hadronic calorimeter (HCAL) main purpose is to measure the energies and the directions of hadrons inside jets, and it is also used to evaluate the missing transverse energy. To achieve these goals, it has to present fine granularity and good hermeticity, covering a solid angle as large as possible. It works as a sampling calorimeter, where the absorbing material is brass, which facilitates the development of hadronic showers, and the active material, which detects the particles produced in the showers, is a plastic scintillator. The photons produced by the scintillators pass through a wavelength shifter and are then read by photodetectors. The HCAL design has been strongly influenced by the limited space available between the outer extent of the ECAL (at radius r = 1.77 m) and the inner extent of the magnet coil (r = 2.95 m) so that in the barrel only 5.82 interaction lengths (λ_0) of absorbing material are present. This is recovered by putting an additional layer of scintillators, called Hadron Outer (HO), just outside the magnet. The HCAL presents the typical barrel-endcaps structure and covers the area $|\eta| < 3$. To enhance the hermeticity, the $|\eta|$ coverage is extended up to 5 by putting two additional Hadron Forward (HF) detectors at a distance of 11 m along z with respect to the interaction vertex. In the HF, the brass and the scintillators are replaced by quartz fibres embedded in steel, in order to resist the much higher radiation dose of the very forward region. An overview of the structure of the HCAL detector is shown in Figure 4.14. The resolution achieved by the HCAL is the following:

$$\frac{\sigma(\mathrm{E})}{\mathrm{E}} = \frac{1}{\sqrt{\mathrm{E/GeV}}} \cdot 85\% \oplus 7.4\% \tag{4.10}$$

where the first term is stochastic and the second is due to the non-uniformity and the calibration uncertainty, as in the case of the ECAL. The HCAL is not as performing as the ECAL due to the different experimental approach (it is a sampling calorimeter, while the ECAL is a homogeneous one), because of the limited amount of absorbing material, and because hadronic showers intrinsically present lower particle multiplicity with respect to electromagnetic ones, leading to wider statistical fluctuations.

4.2.4 The Muon System

Three different detectors are located outside the superconductive solenoid in order to identify muons and measure their momenta. They are based on three different technologies: drift tubes (DT) in the barrel, cathode strip chambers (CSC) in the endcaps, and resistive plate chambers (RPC) both in the barrel and in the endcaps. All the muon chambers are aligned roughly perpendicularly to the muon trajectories and are distributed in order to avoid acceptance holes for values of $|\eta|$ up to 2.4. A view of the muon system, with

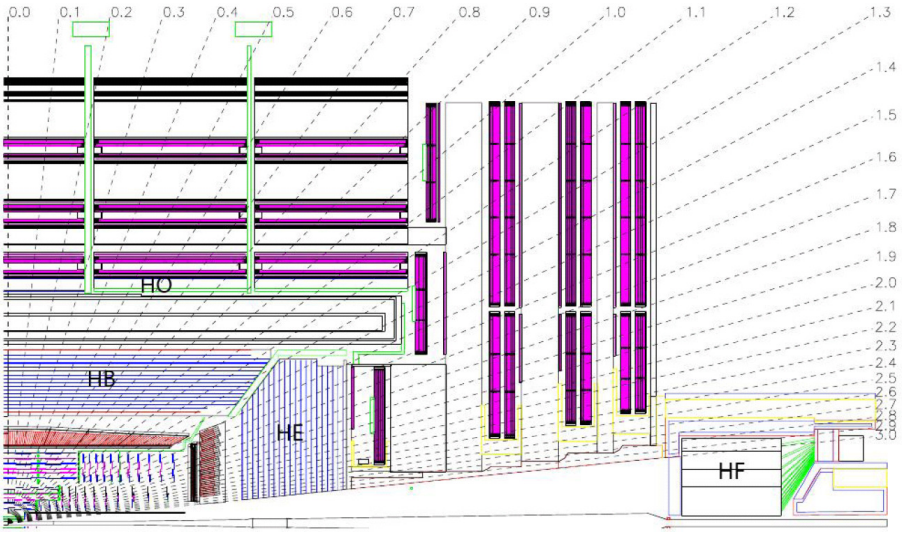
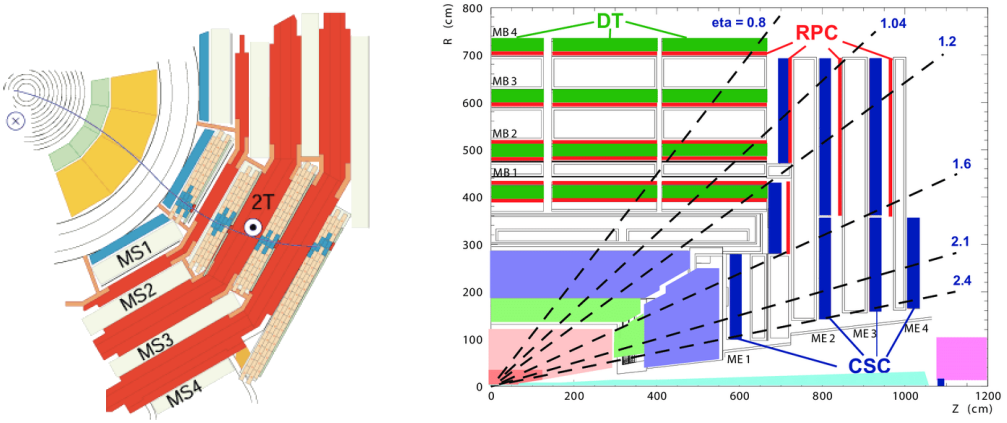


Figure 4.14: Longitudinal view of the CMS HCAL detector in the rz-plane, showing the locations of the different parts of the detector: the barrel (HB), the endcap (HE), the hadron outer (HO), and the hadron forward (HF).



(a) Lateral view.

(b) Longitudinal view.

Figure 4.15: Schematic representation of the CMS muon system. The lateral view on the left (a) shows the typical trajectory of a muon. On the right (b), the view is presented in the rz-plane.

the different detectors individually highlighted, is shown in Figure 4.15. In the barrel, DTs cover the region $0 < |\eta| < 1.3$, where the neutron-induced background is small, the muon flux is low, and the residual magnetic field in the chambers is low. In the endcaps, the CSCs are located at $0.9 < |\eta| < 2.4$, where the neutron-induced background, the muon flux, and the residual magnetic field are high. The RPCs, which cover the region $0 < |\eta| < 2.1$, have a good time resolution, which allows them to identify unambiguously the correct bunch crossing, but present a coarser position resolution than DTs and CSCs. The muon system presents at least 16 layers of material up to $|\eta| = 2.4$. Four stations of detectors are arranged in the cylinders interleaved with the iron yoke in the barrel region, while in each of the endcaps the CSCs and the RPCs are arranged in four disks perpendicular to the beam

and in concentric rings: three rings in the innermost station and two in the others. This allows reaching a muon transverse momentum resolution between 8% and 15% at 10 GeV, and 20% to 40% at 1 TeV. When combining the muon system information with tracker measurements, the resolution is improved to 1.0% to 1.5% at 10 GeV, and from 6% to 17% at 1 TeV. The muon system contains a total of 25000 m^2 of active detector planes and nearly 1 million electronic channels.

4.2.5 The Trigger System

The huge collision rate delivered by the LHC (40 MHz) is much larger than the acquisition rate achievable by the CMS experiment ($\sim$ 1 kHz). A system to select only the most interesting events and save them has thus been implemented and goes under the name of trigger. The trigger is based on two levels. The first one, called Level-1 trigger, is hardware implemented on individual sub-detectors and involves the calorimetry and muon system, as well as some correlation of information between these systems. The Level-1 selects events based on the presence of trigger primitive objects such as photons, electrons, muons, and jets above given transverse energy thresholds. It also employs the global sum of E_T and E_T^{miss}. Reduced-granularity and reduced-resolution data are used to form trigger objects. The Level-1 trigger reduces the rate to 100 kHz. A scheme of how it works is shown in Figure 4.16. The second one is the High-Level Trigger (HLT) and further reduces the rate to about 1 kHz. It is a software and runs on a dedicated farm of commercial PCs. It works following the regional reconstruction on demand so that only the objects in the regions of the detector where the Level-1 found a particle candidate are reconstructed and the uninteresting events can be rejected as soon as possible. The HLT uses information from all the sub-detectors and performs object reconstruction with an algorithm very similar to the one used offline (see Particle Flow Reconstruction in Section 5.1), providing good energy resolution of trigger objects and making the online event reconstruction and selection close to the ones performed in the offline reconstruction and analyses. The HLT trigger is split

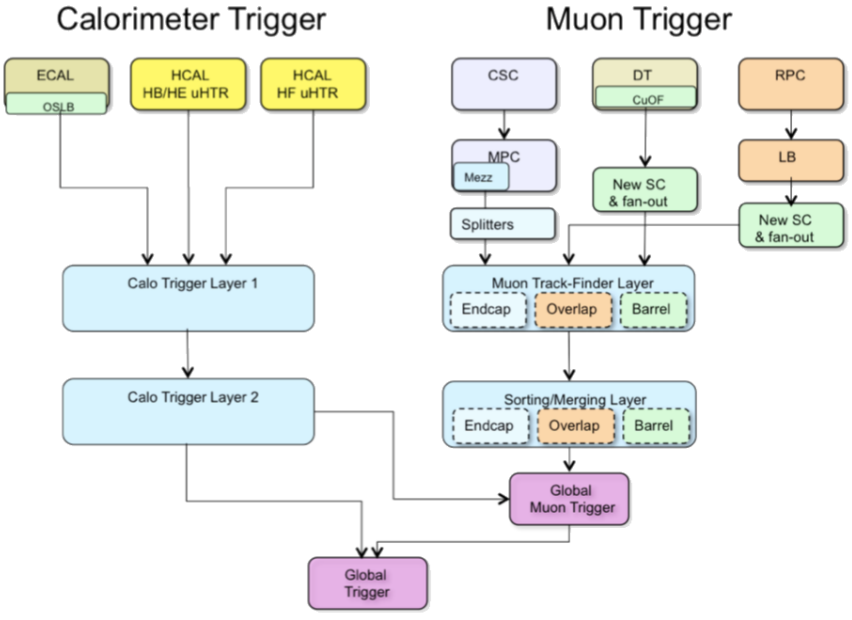


Figure 4.16: Scheme for the Level-1 trigger system showing the main sub-detectors involved.

in *trigger paths*: each trigger path targets a different physics process, so that different trigger paths deal with different objects. A trigger path may require the presence of at least one high-p_T electron, or two muons, or a minimum amount of E_T^{miss} in the event, etc. Trigger paths may be un-prescaled and in this case they save all the events satisfying their requirements, or prescaled and in this case only one good event every N is saved. The need of prescaling a trigger path is due to the very high rate related to it so that by saving all the events would saturate the trigger bandwidth (e.g. trigger paths requiring the presence of very soft leptons). These kinds of triggers are typically used for studies of detectors or specific kinematic regions.

Chapter 5

Physics Objects Reconstruction

Ho un codice segreto
Ho un codice cifrato

CCCP, *Sono come tu mi vuoi*

The different detectors that compose CMS produce electrical signals when particles pass through them. These signals are interpreted as hits in the tracker and in the muon system, or energy deposits in the calorimeters. Specific algorithms are needed to take these raw pieces of information and use them to reconstruct the particles that produced them and their properties (direction, energy and type). They use information from all the CMS subsystems coherently to identify the final-state particles of an event, following the concept of *particle flow* [102]. In this chapter, a description of the general ideas of the particle flow algorithm is presented, followed by a more specific discussion on the reconstruction methods related to the different physics objects, with a particular focus on those used in the analysis presented in this document: electrons, muons, jets and missing transverse energy.

5.1 Particle Flow Reconstruction

The particle flow algorithm [103] is the paradigm for event reconstruction at CMS. It combines all the information from the different CMS sub-detectors in order to reconstruct and identify all the final-state particles in an event, as illustrated in Figure 5.1. These particles are muons, electrons, photons, neutral hadrons and charged hadrons. They are used to build a set of particle-based objects and observables, among them jets, missing transverse energy, lepton isolation and b-jet taggers. The particle flow algorithm follows the scheme presented here:

- **Muons** are reconstructed first and are identified as tracks reconstructed in the silicon tracker associated with a track or several hits in the muon chambers and corresponding to small energy deposits in the calorimeters. The hits in the tracker and in the muon chambers, and the calorimeter deposits associated with the reconstructed muons are removed from the event so that they are not used in the reconstruction of electron and jets, which are the next steps.
- **Electrons** are identified as charged particle tracks associated with one or more ECAL energy deposits, corresponding to the charged tracks and possible bremsstrahlung

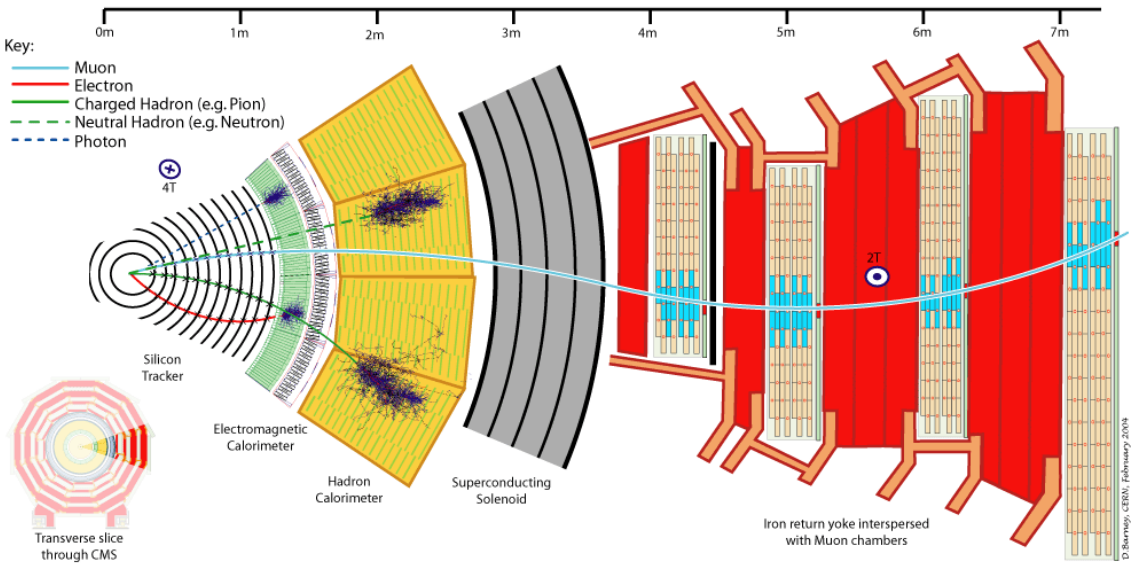


Figure 5.1: View of a section of CMS in the transverse plane. The expected interactions of different particles are shown and illustrate the basic concept of the particle flow reconstruction algorithm.

photons emitted in the tracker material. Also in this case the tracks and the energy deposits in the ECAL used to reconstruct electrons are removed before performing the next steps.

- **Charged hadrons** are identified as charged particles tracks associated with compatible energy deposits in the ECAL and in the HCAL.

- **Photons** are identified as ECAL energy deposits not associated with the extrapolation of any charged particle tracks pointing to the ECAL.

- **Neutral hadrons** are identified as HCAL energy deposits not linked to any charged hadron track, or as ECAL and HCAL energy deposits which exceed the expected energy deposit from a charged hadron.

5.2 Muon Reconstruction

With a lifetime of about 2.2 μs and moving at a speed very close to the speed of light, muons easily escape the CMS detector before decaying, so they are considered as stable particles by the reconstruction algorithms. Since they are charged particles, they interact in the tracker silicon detector and in the specifically-designed muon chambers. On the other hand, muons produced by pp collisions at the LHC typically present momenta which span from hundreds of MeV to hundreds of GeV: in this range of momenta, muons behave as *minimum ionizing particles*, which means that they barely interact in the calorimeters, depositing very small amounts of energy. The muon signature in CMS is thus given by a track in the inner tracking system matched with hits in some of the outer muon chambers and small energy deposits in the calorimeters. The reconstruction of a muon as a particle flow candidate is based on three algorithms [104], which are described in this section.

Standalone Muons

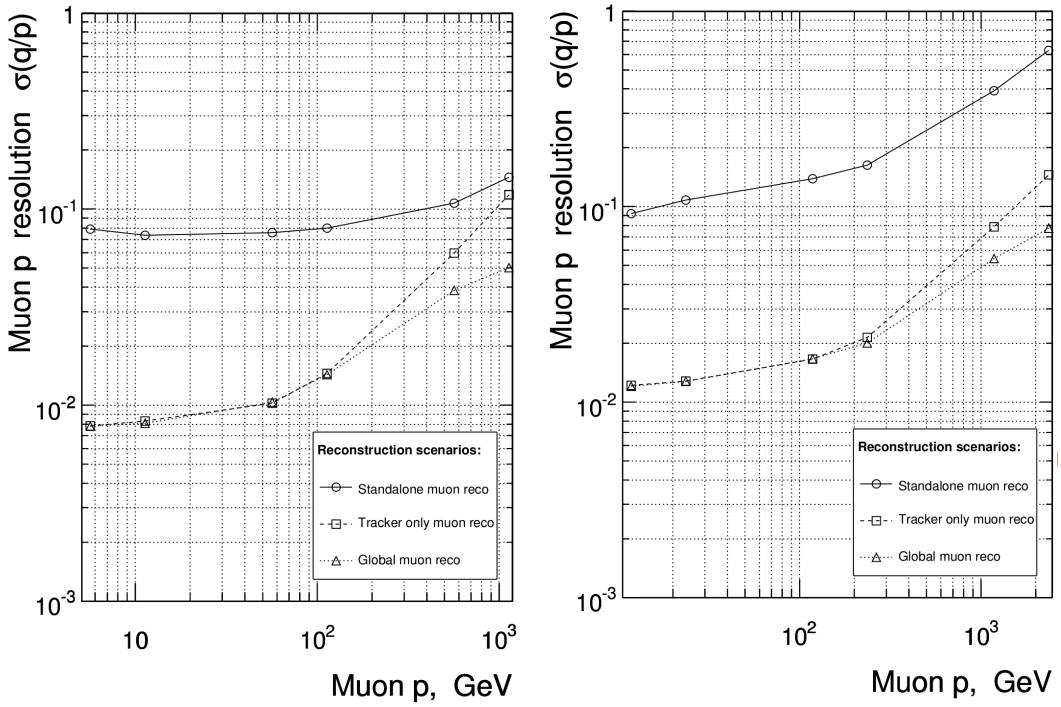
The standalone algorithm uses information from the three muon systems. To reconstruct the trajectory of a muon, it starts by looking for hits in the DT, CSC and RPC subsystems. For each chamber which presents a signal, a segment of the trajectory is built with the hits in the layers inside the chamber, using pattern recognition. A vector of track position, momentum and direction is associated to each segment. A Kalman Filter [105] technique is used to perform a fit to the vectors and reconstruct the muon trajectory. The innermost vectors are taken as seeds of the fit and propagated to the next chamber (*prediction*), taking into account multiple scattering, energy losses in the muon chambers and return yoke, looking for a compatible segment (*measurement*), and the trajectory is updated accordingly [106]. This operation is repeated iteratively until the outer chamber is reached. For each segment added, a χ^2 selection is applied, in order to evaluate how the quality of the fit to the trajectory is affected by including a new chamber and reject bad hits, which may come from showering, delta rays or pair production. Once the trajectory has been built from inside to outside, the same Kalman Filter technique is applied from outside to inside. Once the inner chamber has been reached, the trajectory is extrapolated to the point of closest approach to the beamline, where can be constrained by vertex requirements. To be considered as a standalone muon, the trajectory needs to pass some additional criteria: at least two *measurements* must be present in the fit, one of them coming from DT or CSC chambers, allowing the suppression of fake DT or CSC segments due to combinatories.

Global Muons

Global muons are defined by combining standalone muons tracks with independently reconstructed tracks in the inner tracking system. The matching is performed by extrapolating such tracker tracks to the inner chamber of the muon system and looking for standalone muon tracks compatible in terms of momentum, position and direction. Once two tracks are matched, their hits are used to perform a new global fit. Poor matches between the inner and the standalone tracks are avoided by requiring quality criteria. The global muons thus take advantage of both the tracker and the muon chambers to improve the description of the muon properties. The tracker shows, in fact, better momentum resolution for $p \lesssim 200$ GeV, thanks to the higher position precision and the larger number of hits it can provide. On the other hand, for very large momenta, due to the reduced bending in the magnetic field, the additional hits at high radius given by the muon chambers help to recover the resolution degradation of the silicon tracker, as shown in Figure 5.2.

Tracker Muons

Tracker muons are defined as tracks reconstructed in the silicon tracker, which can be associated with segments in the muon chambers, or to energy deposits in the calorimeters compatible with the muon hypothesis. In particular, the reconstruction algorithm takes into account all tracker tracks with $p_T > 0.5$ GeV and $p > 2.5$ GeV and considers them as tracker muons if they can be matched to at least one muon segment. Additionally, energy deposits compatible with a muon passing through the calorimeters can be used for muon identification. Tracker muons are particularly interesting at low p_T when the momentum resolution of global muons degrades.



(a) Barrel $\eta = 0.5$.

(b) Endcap $\eta = 1.5$.

Figure 5.2: Resolution of muon (1/p) versus p for standalone, global, and tracker-only reconstructions, in barrel (left) and endcap (right).

5.3 Electron Reconstruction

Electrons in CMS interact with the tracker system and the electromagnetic calorimeter. Their signature consists of a single track matched to an ECAL energy deposit (cluster). When an electron reaches the ECAL, it starts to produce an electromagnetic shower almost immediately, and most of the energy is collected by a small number of crystals around the one initially hit. Passing through the tracker layers, the electron loses part of its energy as bremsstrahlung, emitting photons before reaching the calorimeter. As the electron energy goes down, the bending inside the magnetic field increases, producing a spread of the emitted photons along the ϕ coordinate. In order to reconstruct the original electron energy, it is thus essential to recognize which photons are associated with the bremsstrahlung process and account for their energy. This is the purpose of superclustering, the first stage of the electron reconstruction algorithm in CMS, which results in the measurement of the electron energy. After that, the track-building stage and the matching between the tracker and the ECAL information follow.

Supercluster Reconstruction

Superclusters (SCs) in the ECAL are built by grouping together all crystals contiguous to a seed crystal if their energy deposit is two standard deviations above the electronic noise [107]. The requirement of a crystal to be taken as a seed is that its energy must be above a certain energy threshold, depending on η. In the barrel, this energy threshold E_{seed} is 230 MeV. In the endcaps, both the energy and the transverse energy are considered: a crystal to be considered a seed must satisfy $E_{seed} > 600$ MeV, or $E^{T}_{seed} > 150$ MeV.

Electron Track

Once the SC has been found, the reconstruction proceeds with the track-building stage [108]. Considering both the positive and negative charge hypotheses, the position of the supercluster is back-propagated in the magnetic field to the nominal vertex, looking for compatible hits in the pixel detector. If a pair or a triplet of such hits is found, they are used as seeds to build the electron track: the trajectory is reconstructed taking into account energy losses in the material and is fitted with a Gaussian Sum Filter (GSF). This allows approximating the electron energy loss probability density function described by the Bethe-Heitler [109] with a sum of Gaussian functions.

GSF Electron

Finally, the supercluster and the track information are merged together. In this way, the energy measurement E_{SC} is combined with the tracker momentum measurement p_{tk}, as shown in Figure 5.3. This allows to improve the resolution obtained with the two methods separately, mainly due to the opposite behaviour of the ECAL energy resolution and the tracker momentum resolution, and because E_{SC} and p_{tk} are differently affected by the bremsstrahlung radiation. For electrons of $E > 15$ GeV, the solution is driven by the ECAL performance, while for low-energy electrons the momentum resolution is more important.

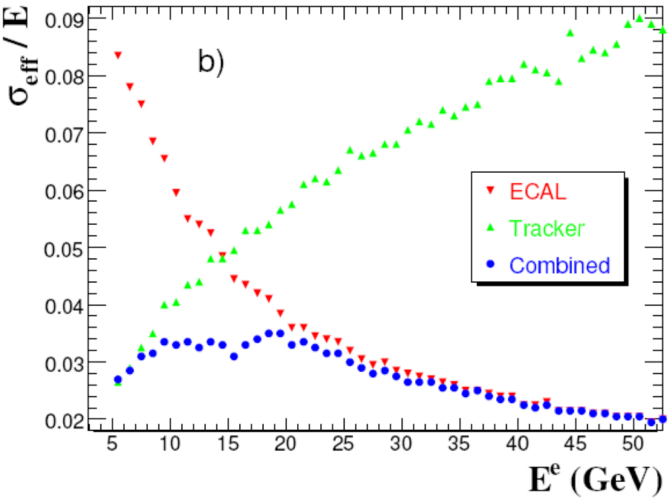


Figure 5.3: Electron energy resolution as a function of energy, measured with ECAL supercluster (red), electron track (green), and the combination of the two (blue) [94].

5.4 Jet Reconstruction

Due to the colour confinement discussed in Section 3.1.2, quarks and gluons are not directly detected in CMS: once they are far enough from the rest of the proton, they start to emit soft and collinear quarks and gluons, until the point where a non-perturbative transition makes the partons to combine into colourless hadrons. This produces a shower of collimated particles, usually called jet. Due to energy conservation, a jet reflects the energy and the direction of propagation of the parton that originated it. From the detector point of view, jets are a set of tracks and energy deposits in a defined region of the experimental

apparatus. Since they are composite objects, an algorithm is needed in order to coherently choose the particles to be considered as proceeding from the original parton. Due to the high probability for a parton to emit a soft or collinear gluon, the chosen algorithm must satisfy some basic requirements, in order to be used to provide a sensitive theoretical prediction. In particular, two conditions have to be satisfied:

- **Collinear safety**: the algorithm output must not change in case a particle of momentum p is substituted by two collinear particles of momentum p/2.
- **Infrared safety**: the algorithm output must not change in case an infinitely soft particle is added or subtracted to the list of the particles to be clustered.

The algorithm that fulfils the previous conditions and is currently used in CMS is called anti-k_T [110]. It introduces the distance d_{ij} between two *entities* (particles or pseudo-jets) still to be clustered, and the distance d_{iB} between the entity i and the beam line, as:

$$d_{ij} = \min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \tag{5.1}$$

$$d_{iB} = \frac{1}{p_{Ti}^2} \tag{5.2}$$

where ΔR_{ij}^2 is the distance between the i and j entities in the $\eta \times \phi$ plane and R is the algorithm radius parameter in the $\eta \times \phi$ plane, usually set by CMS to 0.4. If for two entities the distance d_{ij} is smaller than d_{iB}, they are merged together by summing their 4-momenta. If for an entity i, d_{iB} is smaller than d_{ij}, for all the possible j, i is promoted to a jet and removed from the list of entities. The algorithm is repeated iteratively until only jets are present in the event. The jet momentum is defined as the sum of the momenta of all the particles contained in the jet. In order to have a proper estimation of the energy of the original parton, a set of corrections is applied:

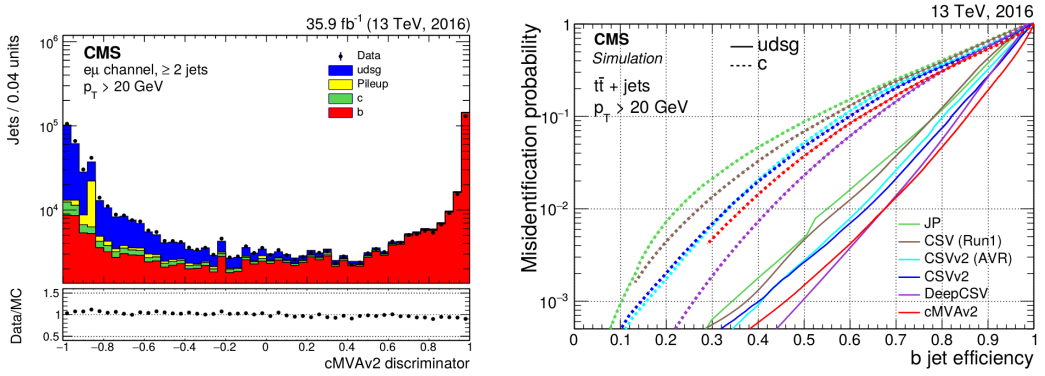
- Level 1 (**offset**): this first step aims to remove, event by event, the contribution to the jet energy given by particles that randomly overlap the jet area, but are produced in secondary proton-proton collisions within the same bunch crossing (pile-up), or by the underlying event.
- Level 2 (**relative**): this second step is required to have the same jet response along η, taking as reference the centre of the barrel.
- Level 3 (**absolute**): in the last step, γ + jet events are used to set the correct jet energy scale. In events with just one jet and one photon, the two particles must have the same energy, so the better energy resolution provided by the ECAL for photons can be exploited to calibrate the energy scale for jets.

Level 1 corrections are applied both to data and Monte Carlo events. Level 2 and Level 3 corrections are derived from simulation and then checked in real data. Possible differences between data and simulated events are taken into account with further corrections to jets in real data.

5.4.1 B-Tagged Jets

Jets originating from the hadronization of bottom quarks can be distinguished from other jets, thanks to the relatively long lifetime of b-quarks and their larger mass with respect

to light quarks. The hadronization products of a *b-jet* are thus expected to come from a secondary, or displaced vertex, with respect to the primary interaction vertex, since the b-quark can travel few millimetres before decaying. On the other hand, b-jets typically present a larger number of particles, due to the high mass of b-quarks. Furthermore, b-quarks can decay to an electron or muon with a probability of around 20%, so that b-jets can present soft leptons. Several algorithms, based on multivariate techniques, have been developed in order to optimize the discrimination of b-jets [111], by combining all these features. In particular for the analysis presented in this document, the cMVAv2 algorithm has been employed. It makes use of the outputs of two other algorithms (CSVv2 and Jet Probability) to enhance the b-tag efficiency and reduce the light quark mistag probability. In Figure 5.4 the distribution of the cMVAv2 b-tagger and the comparison of the performances of different b-taggers are shown.



(a) cMVAv2 discriminator.

(b) Mistag probability.

Figure 5.4: Distribution of the cMVAv2 discriminant in data and simulation (left), and misidentification probability for several b-taggers, with the cMVAv2 performances drawn in red (right).

5.5 Missing Transverse Energy

The protons beams at the LHC carry almost no transverse momentum. Due to momentum conservation, the vectorial sum of the transverse momenta of all the particles of an event is zero:

$$\sum_j \vec{\mathrm{p}}_{\mathrm{T}}^{\,j} = 0 \tag{5.3}$$

On the other hand, several effects can lead to a momentum imbalance in the transverse plane in an event:

- Presence of neutrinos, which due to their very weak interaction with matter escape the experimental apparatus without being detected.
- Particles going to regions outside the acceptance of the detector, or in dead regions, or particles not being reconstructed.
- Finite momentum resolution of the detectors, so that even in events where no neutrinos are present and where all the particles are reconstructed, the momenta do not sum exactly to zero.

- Exotic particles may behave like neutrinos, being electrically neutral and undetectable for CMS. Of particular interest for this document are the DM particle candidates.

For this reason, it is useful to introduce the *missing transverse momentum*, defined as the negative sum of the momenta of all the reconstructed particles in an event:

$$\vec{\mathrm{p}}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{j} \vec{\mathrm{p}}_{\mathrm{T}}^{j} \tag{5.4}$$

It is common to refer to the module of this quantity as *missing transverse energy* ($\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$). All the particle flow candidates of the events are taken into account to define the so-called particle flow $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$:

$$\text{particle flow } \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} = -\left| \sum_{j,PFcand} \vec{\mathrm{E}}_{\mathrm{T}}^{j} \right| \tag{5.5}$$

5.6 Lepton Isolation

The isolation of a lepton measures the amount of activity close to the lepton itself and is an effective variable to distinguish prompt leptons (i.e. coming from the decay of a W boson) from leptons produced by the decay of heavy flavour quarks. In this second case, a large hadronic activity is in fact expected around the leptons, since quarks produce hadronic jets, while W boson leptonic decays do not. Different algorithms for the definition of the isolation exist, but all of them evaluate the amount of activity in a cone in the (η,ϕ) plane around the lepton. Geometrically, the cone is defined by the condition:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{5.6}$$

An illustration of the isolation cone is shown in Figure 5.5. Depending on the isolation definition, energy deposits in the calorimeters or tracks momenta measured in the tracker can be used to estimate the activity around the lepton. In any case, the lepton track or energy deposits are removed from the isolation calculation. The presence of pile-up can bias the isolation measurement since an increased hadronic activity is usually associated with a higher level of pile-up. Charged hadrons can be associated with the primary vertex that produced them thanks to the high position resolution of the tracker, such that only charged hadrons coming from the primary vertex are included in the isolation computation. For neutral hadrons, this is not straightforward so that corrections to isolation are applied to mitigate this effect. In this document, isolation is computed using particle flow candidates, or tracker tracks.

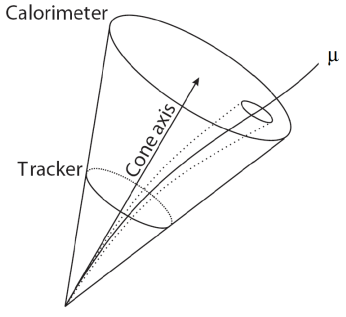


Figure 5.5: Schematic representation of the cone used for the isolation computation.

Chapter 6

Event Simulation

Punta sul nero
Punta sul rosso
Punta di più

CCCP, *Tu menti*

Data have to be compared with theoretical predictions to allow hypothesis falsifiability. For this purpose, the so-called Monte Carlo (MC) event simulation approach is used. It allows generating a large number of simulated events to reproduce the expected kinematic distributions of a process, recreating the physics of the proton-proton scattering, and the interaction of the particles inside the detector. Since differences between data and simulation can occur, in particular when describing the detector, corrections to the MC are put in place. In this chapter, a short description of the Monte Carlo simulation method is provided, followed by an introduction to the method used to correct it for the differences with respect to real data.

6.1 The Monte Carlo Method

To test a physics hypothesis (e.g. the possibility to produce dark matter at the LHC), the outcome of a physics experiment has to be compared with the predictions of one or more models. In the particular case of this document, the SM is taken as the hypothesis to be rejected, in the sense that if DM can be produced at the LHC, the SM alone would not be able to explain the results obtained. Two possible extensions of the SM, described in Chapter 7, represent the hypothesis to be verified. These three models provide predictions in the form of cross sections, or production rates, such that the comparison with the experimental data is not trivial, and can happen only via a statistical interpretation. In order to perform this comparison, events are simulated, according to the cross sections and the kinematic predictions of the different models, using the Monte Carlo method [112]. The basic idea is to implement a random number generator able to reproduce the probability density functions for several observable quantities provided by the equation describing a certain process, in order to simulate a large number of events. This task is performed by specific software, called *event generators*. These programs operate several steps in order to provide a reliable description of the simulated event. The first step consists in the simulation of the *hard scattering* between partons for the selected process; here the decay of unstable particles inside the detector volume is also produced. After that, the

underlying event, produced by secondary interactions between the remnants of the incoming hadrons, and the *parton showering*, involving incoming and outgoing coloured particles, are simulated. Finally, the consequent *hadronization*, that converts the showers into outgoing hadrons is performed. The crossing of all the final-state particles produced so far through the detector must then be simulated, such that hits in the tracking system and calorimeter energy deposits provide an event format that allows the comparison with real data. In this step, the interaction of the particles with the detector material, which can cause multiple scattering or energy losses, is taken into account. The standard packages used to simulate the interaction of the particles with the detector are GEANT4 [113], for a full and detailed description of the interactions of the particles in the material, and DELPHES [114], which provides a parametrized, and hence faster simulation. In this Section, the main steps of the generation process are described.

6.1.1 Hard Scattering

The hard scattering, involving large momentum transfer between the colliding particles, is the key process to produce events of interest at the LHC, where heavy mediator particles and energetic particles in the final state are involved. It is produced by the collision of two partons inside the protons, and is described by the parton-level cross section $\hat{\sigma}_{\mathrm{ab}\to\mathrm{n}}$, where a,b are the incoming partons and n denotes the number of outgoing particles. This quantity is related to the hadronic cross section $\sigma_{\mathrm{h_1h_2}\to\mathrm{n}}$, where $\mathrm{h_1}$,$\mathrm{h_2}$ are the incoming hadrons, in this case protons, through the Parton Distribution Functions (PDFs), according to the following equation:

$$\sigma_{pp\to n} = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a^{h1}(x_a,\mu_F) f_b^{h2}(x_b,\mu_F) d\hat{\sigma}_{ab\to n}(\mu_F,\mu_R) \tag{6.1}$$

Where $f_a^{h1}(x_a,\mu_F)$ and $f_a^{h2}(x_b,\mu_F)$ are the PDFs of the two protons, which depend on the momentum fraction x and on the factorization scale μ_F [115], and μ_R is the renormalization scale [115]. The parton-level cross-section, on the other hand, can be written as the product of a matrix element squared, $|\mathcal{M}_{ab\to n}|^2$, which can be evaluated using Feynman diagrams, and the final-state phase space Φ_n, such that Eq. 6.1 can be written as:

$$\sigma_{pp\to n} = \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h1}(x_a,\mu_F) f_b^{h2}(x_b,\mu_F) \frac{1}{2\hat{s}} |\mathcal{M}_{ab\to n}|^2(\Phi_n;\mu_F,\mu_R) \tag{6.2}$$

In some sense, this second way of writing the cross section is closer to the one used by event generators, since they provide comprehensive lists of matrix elements and the corresponding phase space parametrization, considering final states with $n = 1$, 2 or 3 particles. Different generators have been used in this document; some of them, like MADGRAPH [116], can compute matrix elements at leading order (LO), while others, as POWHEG [117] and aMC@NLO [118] can reach next-to-leading order (NLO).

6.1.2 Parton Showering

Coloured particles cannot be found alone (see Section 3.1.2), instead they use part of their energy to produce showers of additional coloured particles until the energy becomes too low, and the particles produced in this way combine among themselves to form colourless hadrons (see Section 6.1.4 in the following of this Chapter). The parton showering consists in simulating the processes of shower production for final-state particles produced by the

hard scattering, but also for particles in the initial state and for the remnants of the colliding protons. In fact, initial-state partons can emit a gluon as initial state radiation (ISR), as well as the remnants of the protons can emit a gluon as final state radiation (FSR), and in both cases, a new hadronic shower is produced. The parton showering is usually simulated by using PYTHIA [119] and Herwig [120, 121].

6.1.3 Underlying Event

The underlying event is produced by secondary interactions between the protons remnants, apart from the hard scattering (see Section 4.1.3). Since these secondary interactions typically involve small momentum exchange, they hardly add detectable jets to the event, but in any case increase the number of particles produced at the hadronization step.

6.1.4 Hadronization

The process of binding of coloured partons into colourless hadrons is simulated phenomenologically, since at low energies (hundreds of MeV), the QCD ceases to be a perturbative theory. The event generators usually use the Lund string model [122], in which quarks are bound together via a gluon string. If two quarks tend to travel away one from another, the string gets stretched and accumulates energy. When the energy stored reaches the threshold needed to produce a new $q\bar{q}$ pair, the string breaks and two new quarks appear. The process is repeated until the energy available is below a certain threshold.

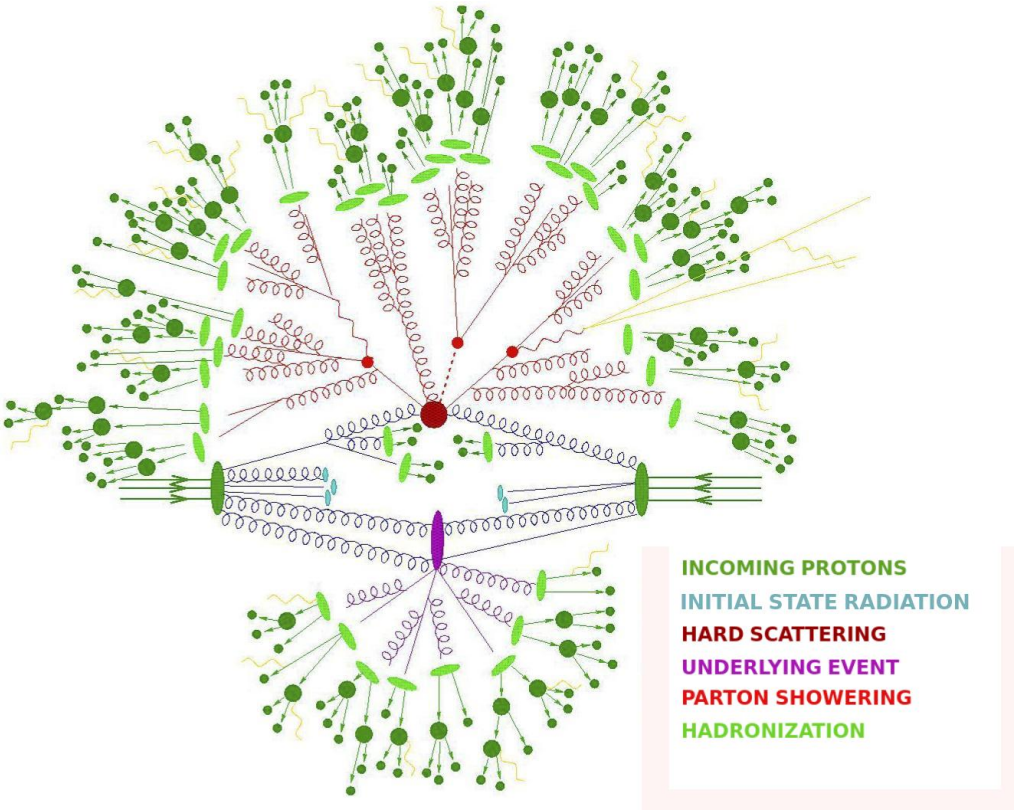


Figure 6.1: Schematic representation of a proton-proton collision. The three green arrows represent the partons inside the colliding protons, the dark red circle the hard scattering, and the violet oval the underlying-event secondary vertex.

6.1.5 Pile-Up

The contribution of pile-up (see Section 4.1.3) to the hard scattering process is reproduced in simulation by superimposing a number of simulated soft interactions vertices to every event. The number of additional vertices is chosen in order to agree with the multiplicity distribution observed in data.

6.2 Monte Carlo Corrections

Several effects can spoil the reconstruction of physics objects in CMS. The sub-detectors present blind or dead zones or have cracks, and the triggering algorithms can use only partially the available information, due to time and storage requirements. These effects generate inefficiencies in the trigger selections and on the reconstruction and identification of physics objects. They cannot be perfectly reproduced by MC, so that differences between data and simulated events may occur. Efficiencies are measured independently in data and MC, and the ratio of the two values is used to correct the simulated events.

6.2.1 The Tag and Probe Method

As mentioned, some of the particles produced in proton-proton collisions may go outside the acceptance region of the detector, and the reconstruction algorithm may fail in reconstructing some of the particles passing through the detector. Since several variables can cause these losses, the estimation of the inefficiency is tackled by a probabilistic perspective, regardless of what produced it during the detection and reconstruction chain. A fully efficient system would be able to detect and reconstruct all the particles produced so that the efficiency can be in principle defined as the fraction of particles reconstructed with respect to the total amount of particles produced. Unfortunately, this last piece of information is not accessible since it is not possible to know how many particles have passed undetected. A technique able to avoid this issue goes under the name of *Tag and Probe* method [123] and is currently used to evaluate efficiencies. It makes use of di-object resonances, as the J/Ψ and the Z, to measure the reconstruction efficiency of the objects in which they decay. Let's assume we want to measure the reconstruction efficiency for muons using the Z resonance. The idea of the Tag and Probe method is to select di-muons events in which the di-muons invariant mass is compatible with the mass of the Z boson. In these events, one muon (*tag*) must pass *Tight* selection criteria (e.g. it must be reconstructed as a global muon), to ensure it is a real muon, thus *tagging* the event as really produced by the decay of a Z boson. The other muon represents instead the *probe* since it tests the efficiency of the reconstruction selections to be measured, typically looser than the *tag* ones (e.g. standalone muon). In fact, if this second muon passes these *Loose* selections, it is counted as a *passing* probe, if it does not pass them, it is counted as a *failing* probe, such that in principle the efficiency ε can be measured as:

$$\varepsilon = \frac{\mathrm{N}_{\text{passing probes}}}{\mathrm{N}_{\text{probes}}} = \frac{\mathrm{N}_{\text{passing probes}}}{\mathrm{N}_{\text{passing probes}} + \mathrm{N}_{\text{failing probes}}} \tag{6.3}$$

Equation 6.3 is not directly employed since it does not take into account possible background contamination from non-resonant processes that accidentally present an invariant mass compatible with the mass of the Z boson. Instead, a signal + background model fit to the di-muon invariant mass is exploited, separately for the passing and the failing probes, such that the non-resonant component can be measured and subtracted from the efficiency computation, as illustrated in Figure 6.2. The fit is performed in different bins of $\mathrm{p_T}$ and

η of the probe, to provide efficiencies as a function of these two kinematic variables. The efficiencies are measured both in data and in simulations. Differences between the actual detector configuration and simulations are taken into account by computing the so-called *scale factors*, defined as the ratio between efficiencies measured in simulation and in real data. The scale factors are then applied to MC events to correct those effects.

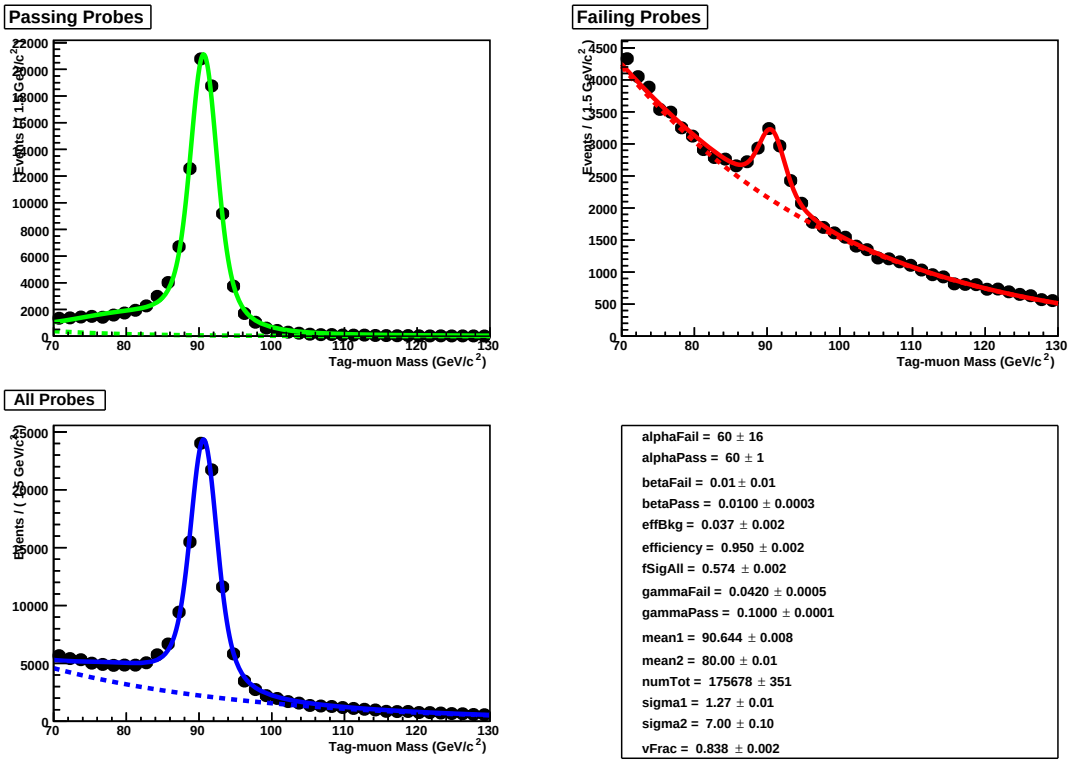


Figure 6.2: Example of the Tag and Probe method, used to evaluate the reconstruction efficiency of muons during the 2016 data taking. The invariant mass distributions of passing and failing probe events and of all the events are fitted separately by using a function able to describe simultaneously the signal and the background distributions.

Chapter 7

Mono-Higgs Physics Models and Main Backgrounds

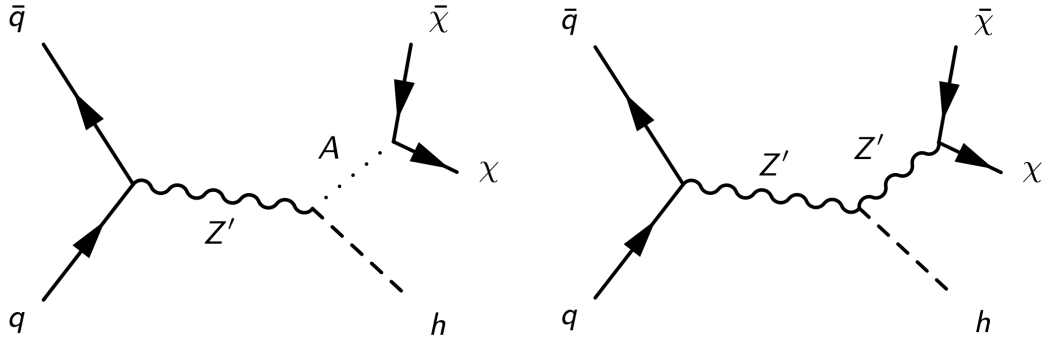
Eri così carino
Pigro di testa
E ben vestito

CCCP, *Tu menti*

The choice of the models to use for this search follows the recommendations of the ATLAS-CMS Dark Matter Forum [7]. It gathers the Mono-X models emerged as most interesting for searches at the LHC after discussions between experimental and theoretical physicists. In this document, two of such *simplified models* are used as benchmarks, which produce a final state with an SM Higgs boson and a pair of DM WIMPs (mono-Higgs). The selected decay channel of the Higgs boson is to a pair of W bosons, both decaying to a lepton and the corresponding neutrino, where the two leptons are of different flavour. The branching fraction BR(h→WW*) is the second largest (21.5% [81]) only after the one to a pair of b-quarks. The choice of the fully leptonic decay of the W bosons partly reduces the statistics of the channel, but ensures a clean final state, with relatively small background contamination and good control of systematic uncertainties. For these reasons, this decay channel can give a contribution to the sensitivity of the mono-Higgs search and is a fundamental part of the plan to have all the possible Higgs boson decay channels inspected. The two benchmark models considered for the search are the Z'-2HDM [124] model and the Baryonic-Z' [6] model and are schematically represented at *tree level* in Figure 7.1. Several SM processes can mimic the signature of the signal, which consists of one muon and one electron of different charge and large E_T^{miss}. This chapter is dedicated to the description of the signal models and to the review of the main backgrounds that can affect this search.

7.1 Z'-2HDM Model

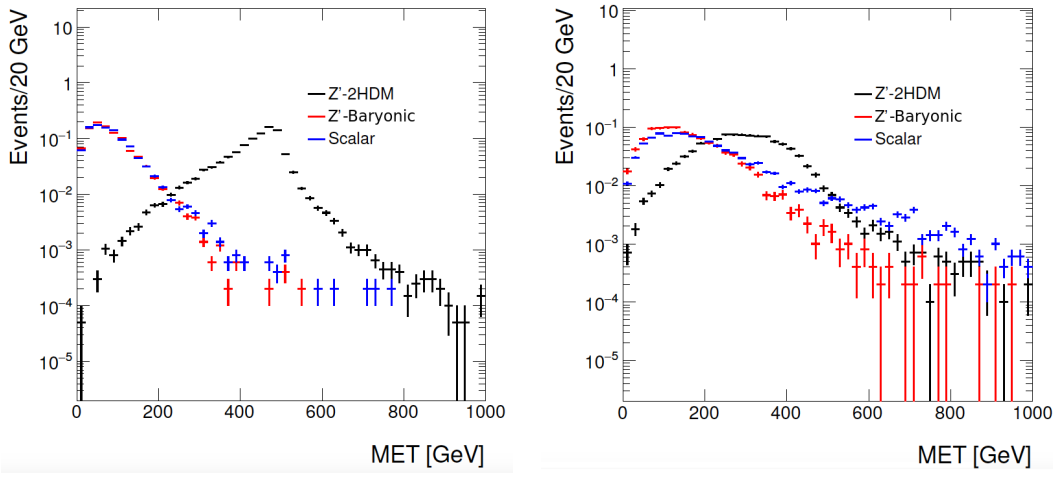
The first model inspected is a Z'-two-Higgs-doublet-model [124]. Here a Z' vector boson is produced by quark-antiquark annihilation and resonantly decays to an SM Higgs boson and a heavy pseudoscalar A, which then decays to a pair of DM particles. A Feynman diagram representing this model at LO is shown in the left part of Figure 7.1. By coupling DM to the pseudoscalar A, it is possible to avoid the strong constraints put by other



(a) Z'-2HDM. (b) Baryonic-Z'.

Figure 7.1: Feynman diagrams of the dark matter simplified models inspected: the Z'-2HDM model on the left and the Baryonic-Z' on the right.

channels, or by direct detection experiments. Furthermore, given the resonant decay of the Z', the kinematics associated with this model can be significantly different from Standard Model processes and from other DM production mechanisms, as illustrated in Figure 7.2. In particular, for heavy mediators the missing transverse energy and the Higgs transverse



(a) Low mass mediators (b) High mass mediators

Figure 7.2: E_T^{miss} distributions for different DM simplified models at generator level [7]. The Z'-2HDM model spectrum peaks at higher value.

momentum are much harder. This model states the existence of two Higgs doublets, of which one couples to up quarks (Φ_u) and one couples to down quarks and leptons (Φ_d):

$$\mathcal{L} \supset y_u Q \Phi_u \bar{u} + y_d Q \Phi_d \bar{d} + y_d L \Phi_d \bar{e} + h.c. \tag{7.1}$$

After electroweak symmetry breaking, the two Higgs doublets get vacuum expectation values v_u and v_d, and can be expressed as:

$$\Phi_d = \frac{1}{\sqrt{2}} \begin{pmatrix} -H^+ \sin\beta \\ v_d + h\sin\alpha + H\cos\alpha - iA\sin\beta \end{pmatrix} \tag{7.2}$$

$$\Phi_u = \frac{1}{\sqrt{2}} \begin{pmatrix} H^+ \cos\beta \\ v_u + h\cos\alpha + H\sin\alpha + iA\cos\beta \end{pmatrix} \tag{7.3}$$

here h and H are neutral CP-even scalars, $H^{\pm}$ is a charged scalar and A is the already mentioned CP-odd scalar. The β angle is defined by the relation $\tan\beta = \dfrac{v_u}{v_d}$, and α as the angle that diagonalizes the h - M mass squared matrix. If h is considered as the SM Higgs boson, then it is valid that $\alpha = \beta - \pi/2$ and that $\tan\beta \geq 0.3$.

7.1.1 Parameter Scan

The model is described by the following parameters:

- The pseudoscalar mass $\mathrm{m_A}$
- The DM particle mass m_χ
- The Z' mass $\mathrm{m}_{Z'}$
- $\tan\beta = \dfrac{v_u}{v_d}$
- The Z' coupling strength g_Z

To study how the kinematics depends on these parameters, simulated signal samples have been produced using MadGraph5_aMC@NLO for the matrix element, PYTHIA 8 for the parton shower and DELPHES for a parametrized interaction with the detector. As shown in Figure 7.3, no kinematic dependence on the particular choice of $\tan\beta$ has been observed, and the production cross section simply scales as a function of $\tan\beta$. Following the recommendation of the ATLAS-CMS Dark Matter Forum, the value of $\tan\beta$ has been fixed to 1. Similarly, the value of g_Z does not affect the kinematic distributions for this model and has been fixed to 0.8, which is the maximum value allowed by electroweak global fits and dijet constraints, as described in [124]. Since the DM particles are produced by the decay of the pseudoscalar A, no significant changes in kinematics is expected by varying the mass m_χ, as long as A is produced on-shell ($\mathrm{m}_\chi < \mathrm{m_A}/2$), as shown in Figure 7.4. The mass of the DM particles has been fixed to $\mathrm{m}_\chi = 100$ GeV. On the other hand, significant dependence on the Z' and A masses has been observed, as shown in Figures 7.5 and 7.6. For this model a scan on these two parameters has been performed, considering masses of the Z' mediator between 600 GeV and 2500 GeV, and masses of the pseudoscalar A between 300 GeV and 800 GeV. A summary of the mass points considered in this document and the corresponding cross-sections is presented in Table 7.1.

	$\mathrm{m}_{Z'}$ [GeV]							
$\mathrm{m_A}$ [GeV]	600	800	1000	1200	1400	1700	2000	2500
800			6.37×10^{-5}	1.31×10^{-4}	1.28×10^{-4}	8.53×10^{-5}	4.92×10^{-5}	1.85×10^{-5}
700			1.98×10^{-4}	2.32×10^{-4}	1.85×10^{-4}	1.07×10^{-4}	5.79×10^{-5}	2.06×10^{-5}
600		2.32×10^{-4}	4.38×10^{-4}	3.65×10^{-4}	2.52×10^{-4}	1.31×10^{-4}	6.73×10^{-5}	2.29×10^{-5}
500		8.43×10^{-4}	7.92×10^{-4}	5.31×10^{-4}	3.30×10^{-4}	1.59×10^{-4}	7.85×10^{-5}	2.57×10^{-5}
400	1.47×10^{-3}	2.09×10^{-3}	1.36×10^{-3}	7.91×10^{-4}	4.56×10^{-4}	2.07×10^{-4}	9.85×10^{-5}	3.12×10^{-5}
300	1.03×10^{-2}	6.39×10^{-3}	3.31×10^{-3}	1.75×10^{-3}	9.58×10^{-4}	4.16×10^{-4}	1.93×10^{-4}	6.00×10^{-5}

Table 7.1: Summary of cross sections in pb for different mass points inspected for Z'-2HDM model. Cross sections are calculated considering $g_Z = 0.8$, $\tan\beta = 1$ and $\mathrm{m}_\chi = 100$ GeV. The branching ratio of $\mathrm{h} \to \mathrm{W^+W^-} \to \ell^+\bar{\nu}\ell^-\nu$ is included in the calculation.

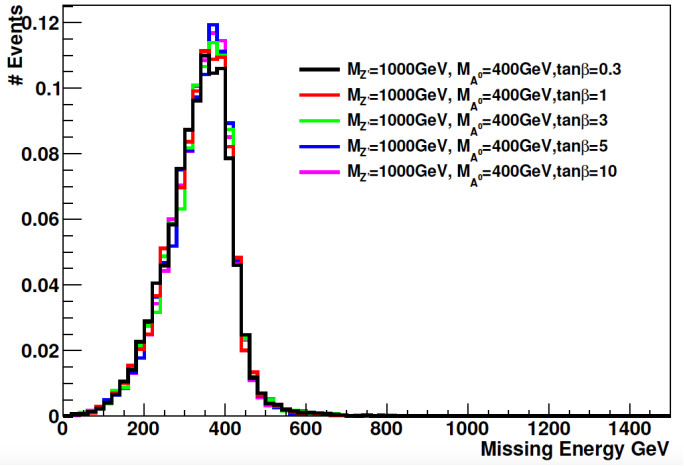


Figure 7.3: Missing transverse energy distributions for Z'-2HDM model varying the tan β parameter. No significant dependence has been observed [7].

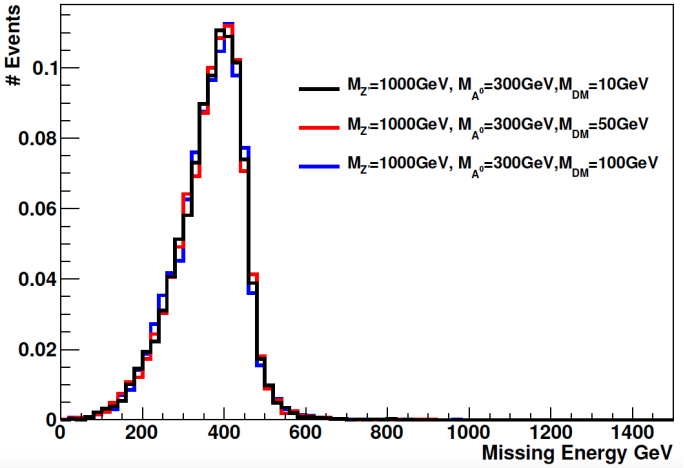


Figure 7.4: Missing transverse energy distributions for Z'-2HDM model varying the mass of the DM particle. No significant dependence has been observed [7].

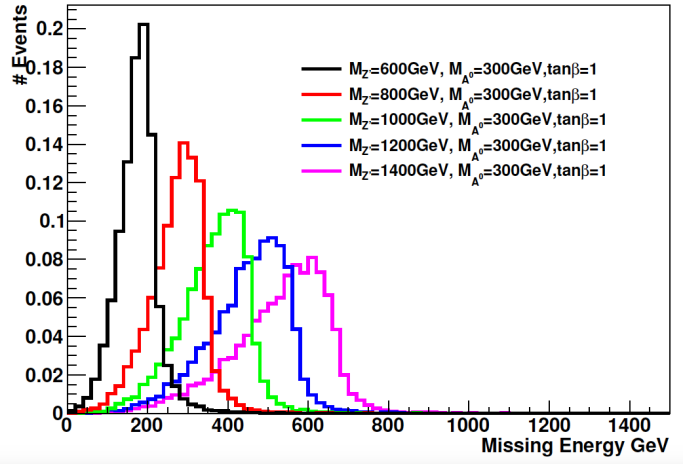


Figure 7.5: Missing transverse energy distributions for Z'-2HDM model varying the mass of the Z' mediator [7].

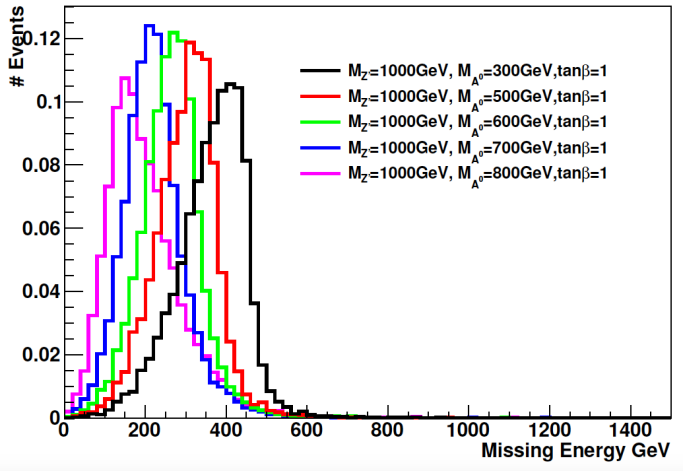


Figure 7.6: Missing transverse energy distributions for Z'-2HDM model varying the mass of the heavy pseudoscalar A [7].

7.2 Baryonic-Z' Model

The second model studied in this document introduces a new vector gauge boson Z', which can irradiate an SM Higgs boson and then decay to a pair of DM particles [6]. A schematic representation at the LO of this model is illustrated on the right side of Figure 7.1. The Z' in this case comes from a new $U(1)_B$ baryon number symmetry, and the DM particles correspond to the stable baryonic states of the model. A baryonic Higgs boson h_B is introduced to give mass to the Z' and mixes with the SM Higgs boson. The Z' mediator interacts with quarks and DM particles, according to the following Lagrangian:

$$\mathcal{L} = g_q \bar{q} \gamma^\mu q Z'_\mu + g_\chi \bar{\chi} \gamma^\mu \chi Z'_\mu \tag{7.4}$$

Here the quark coupling g_q to the Z' is fixed to 0.25 of the gauge coupling g_B, while the coupling of the Z' with the DM particles g_χ is proportional both to the baryon number and the gauge coupling: $g_\chi = B g_B$. No coupling between Z' and leptons is allowed, in order to avoid the constraints coming from the dilepton analyses. When the mixing between the baryonic and the SM Higgs bosons is considered, at energies below $m_{Z'}$ the model is described by the following Lagrangian:

$$\mathcal{L}_{\text{eff}} = \frac{g_q g_\chi}{m^2_{Z'}} \bar{q} \gamma^\mu q g_\chi \bar{\chi} \gamma^\mu \chi \left(1 + \frac{g_{hZ'Z'}}{m^2_{Z'}} h\right) \tag{7.5}$$

Despite the Lagrangian in Eq. 7.5 is just an effective one, it highlights the interactions of the Z' mediator. In any case, the event generation uses the full dimension-4 operator. The first term of the Lagrangian describes the production of the Z' mediator and its decay to a pair of DM particles, while the second term introduces interactions between Z' and the SM Higgs boson, through the coupling $g_{hZ'Z'} = \frac{m_{Z'} 2 \sin\theta}{v_B}$, where θ is the mixing angle between the SM Higgs and the baryonic Higgs and v_B is the Baryonic Higgs vacuum expectation value.

7.2.1 Parameter Scan

The model is described by the following set of parameters:

- The Z' mediator mass $m_{Z'}$
- The DM particle mass m_χ
- The coupling between Z' and DM particles g_χ
- The coupling between Z' and quarks g_q
- The mixing angle between the SM Higgs and the baryonic Higgs $\sin\theta$
- The coupling between the SM Higgs and the Z' mediator $g_{hZ'Z'}$

Also in this case the kinematic dependence on the different parameters has been investigated through the production of signal samples in which the parameters are varied one at a time. As shown in Figure 7.7, the kinematics is not affected by the particular choice of the couplings of the Z' mediator between the DM particles and between the SM Higgs boson. The same is true for g_q and $\sin\theta$. These parameters can only change the value of the cross-sections of the model and are bound by perturbative requirements:

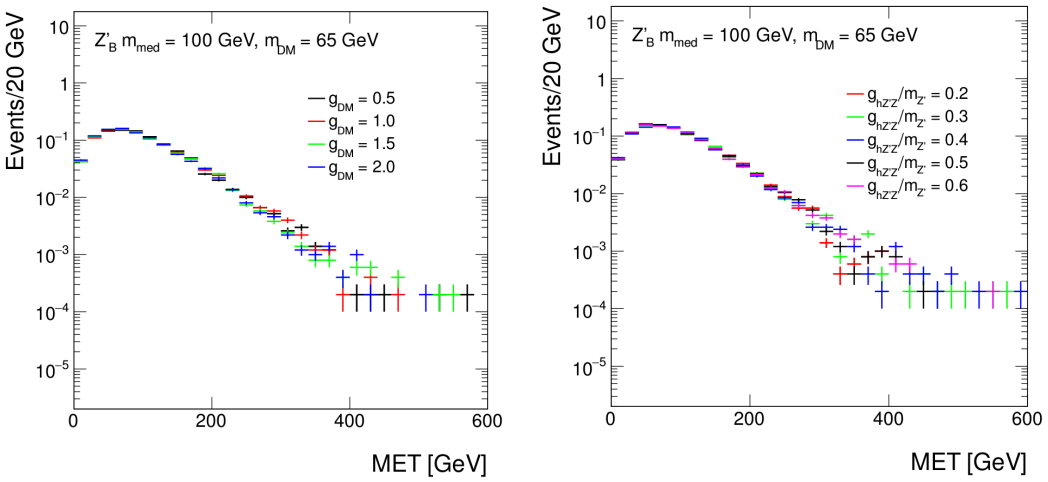
$$g_q, g_\chi < 4\pi \tag{7.6}$$

$$g_{hZ'Z'} < \sqrt{4\pi} m_{Z'} \sin\theta \tag{7.7}$$

According to these constraints and to the recommendations of the ATLAS-CMS Dark Matter Forum, the g_χ parameter has been fixed to 1 and g_q to 0.25. The value $g_{hZ'Z'}/m_{Z'} = 1$ is chosen as it maximizes the cross-section without violating the already mentioned bounds. In Figure 7.8 the dependence of the E_T^{miss} distributions on the choice of the mass of the DM particle is shown, in the case of light or heavy Z' mediators. Since the masses $m_{Z'}$ and m_χ affect the kinematics of this model, a scan on these two parameters has been performed in the analysis. In particular, masses of the Z' mediator between 100 GeV and 2000 GeV have been considered, and masses of the DM particle between 1 GeV and 1000 GeV have been taken into account. Masses of the Z' mediator lower than 100 GeV have not been inspected since in this case the h→Z'Z' channel would be accessible with a non-negligible branching ratio. In Table 7.2 a summary of the mass points used in this document and the corresponding cross-sections are presented.

	m_χ [GeV]					
	1	10	50	150	500	1000
$m_{Z'}$ [GeV]						
2000	3.13×10^{-4}				2.84×10^{-4}	
1995						1.64×10^{-5}
1000	4.53×10^{-3}			4.40×10^{-3}		1.24×10^{-8}
995					2.49×10^{-4}	
500	2.45×10^{-2}			1.52×10^{-2}	4.61×10^{-7}	
300	5.09×10^{-2}		4.37×10^{-2}			
295				3.00×10^{-3}		
200	5.72×10^{-2}		4.69×10^{-2}	7.15×10^{-5}		
100	7.13×10^{-2}	7.12×10^{-2}				

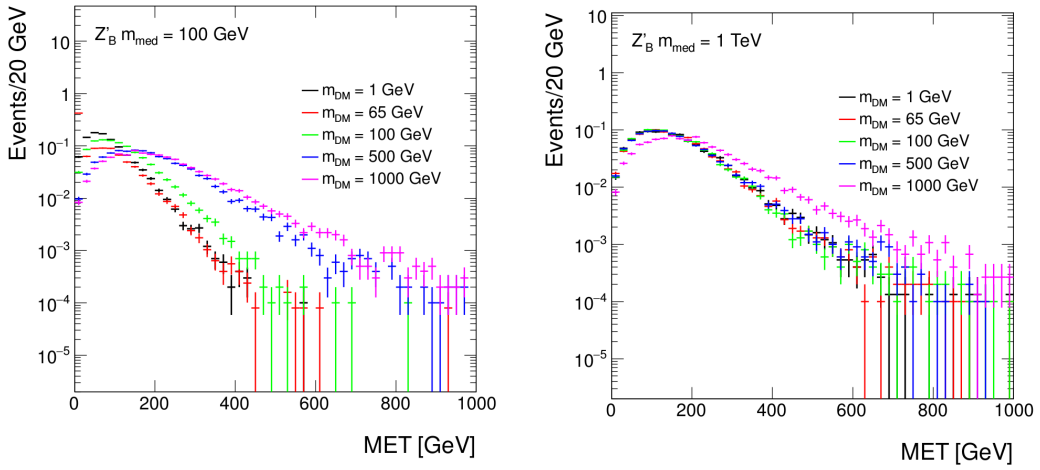
Table 7.2: Summary of cross sections in pb for different mass points inspected for baryonic-Z' model. Cross sections are calculated considering $g_\chi = 1$, $g_q = 0.25$, and $\frac{g_{hZ'Z'}}{m_{Z'}} = 1$. The branching ratio of $h \to W^+W^- \to \ell^+\bar{\nu}\ell^-\nu$ is included in the calculation.



(a) E_T^{miss} vs g_{DM}.

(b) E_T^{miss} vs $g_{hZ'Z'}/m_{Z'}$.

Figure 7.7: Missing transverse energy distributions for Baryonic-Z' model varying the coupling between the Z' mediator and the DM particle (left) and varying the coupling between the Z' mediator and the Higgs boson (right). No significant dependence has been observed [7].



(a) $m_{Z'} = 100$ GeV.

(b) $m_{Z'} = 1000$ GeV.

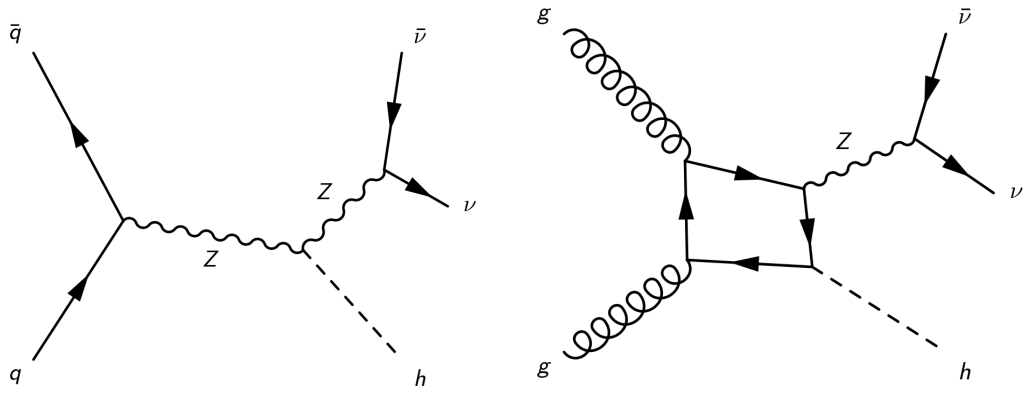
Figure 7.8: Missing transverse energy distributions for Baryonic-Z' model varying the mass of the DM particle. On the left, the mass of the Z' mediator mass is fixed to 100 GeV, while on the right it is fixed to 1000 GeV [7].

7.3 Main Backgrounds

Various SM processes are able to produce a final state similar to the signal one, with two hard leptons of opposite charge and different flavour, and a significant amount of E_T^{miss}. Obviously, the SM production of a Higgs boson decaying to a pair of W bosons is among them, as more in general the production of pairs of W bosons. The fully leptonic decay of $t\bar{t}$ pairs or the production of a top quark and a W boson can be distinguished from the signal since they typically present b-jets in the final state, but due to the high cross section represent an important source of background. The DY process produces a final state similar to the signal one if the Z or γ^* decay to a pair of τ leptons, which in turn decay leptonically, producing an electron, a muon, and four neutrinos, which create a large amount of E_T^{miss}. Also events with one prompt lepton and one lepton produced inside a jet, if the second lepton is wrongly identified as a prompt lepton, can mimic the signal signature. Other processes, typically di-boson or tri-boson production, represent backgrounds if they produce at least two leptons in the final state and one or more additional leptons go in a region out of the detector acceptance or are not reconstructed.

7.3.1 SM Higgs Production

The production of an SM Higgs boson in association with a Z boson, when the Z boson decays to a pair of neutrinos, represents an irreducible source of background for the signal. As shown in the left part of Figure 7.9, the Feynman diagram at LO, when the process is produced by the annihilation of a pair of quarks, is analogous to the signal ones, with the difference that the Z' and the A mediators are replaced by the SM Z boson, and instead of the DM particles χ, the invisible particles are now only neutrinos. The production can also be initiated by the fusion of a pair of gluons, and in this case the final state is obtained through a fermionic box. The cross-sections for the Zh process are very small so that it does not represent a significant source of background after all the selections to define a signal-enriched phase space are applied. In particular, including the branching fraction $h \to W^+W^- \to \ell^-\bar{\nu}\ell^+\nu$ [125], the cross section for the qq→Zh process is 0.017 pb [89], while for the gg→Zh process is 0.00275 pb [89]. Also the other SM Higgs production mechanisms



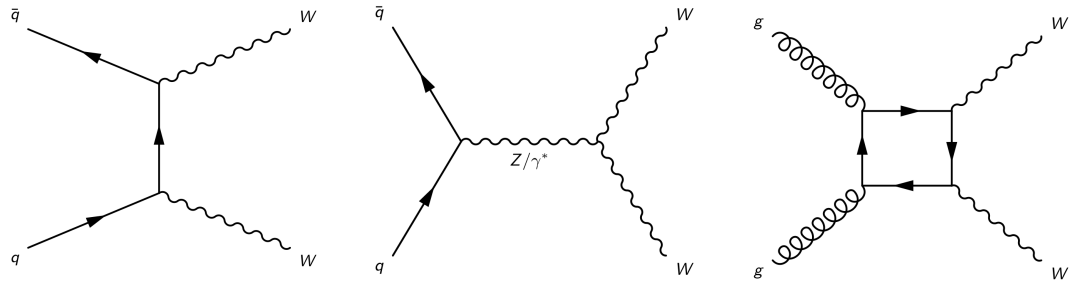
(a) qq→Zh. (b) gg→Zh.

Figure 7.9: Representation of the irreducible background of the analysis: associate production of an SM Higgs boson and an invisibly decaying Z boson. The quark-annihilation production is shown in the left, the gluon-gluon fusion on the right.

are considered as backgrounds, since they present in the final state two prompt leptons and missing transverse energy due to the presence of the neutrinos from the decay of the W bosons. In this case the di-lepton invariant mass distribution is still related to the decay of the Higgs boson, so it is expected to mimic the signal one, but the missing energy module and its angular distribution are not the same as the signal, since here the only contribution comes from the leptonic decay of the W bosons. Several production mechanisms are considered, including the gluon-gluon fusion (ggF) and the vector boson fusion (VBF) already shown in Section 3.2, in addition to the bb→h and the Wh associate production. The cross-sections, considering the fully leptonic decay of the Higgs to WW, are 0.9913 pb for ggF [89], 0.0846 pb for VBF [89], 0.0081 pb for bb→h, and 0.031 pb in the case of Wh associated production [89]. For all the SM Higgs production mechanisms, also the decay to a pair of τ leptons is considered, since the final state in which both the taus decay leptonically can mimic the signal signature. After all the signal selections are applied, the Higgs represents about 5% of the total amount of background.

7.3.2 WW Production

The SM production of a pair of W bosons is the main background of this search. It can mimic the signal signature when the di-lepton invariant mass accidentally reproduces the signal one. Even if this happens only for a small fraction of events, the large cross-section of this process makes it difficult to get rid of it. At the LHC, both quark-quark annihilation and gluon-gluon fusion can generate pairs of W bosons, as schematically shown in Figure 7.10. The largest contribution is due to the quark-quark annihilation channel, for



(a) qq→WW. (b) qq→WW. (c) gg→WW.

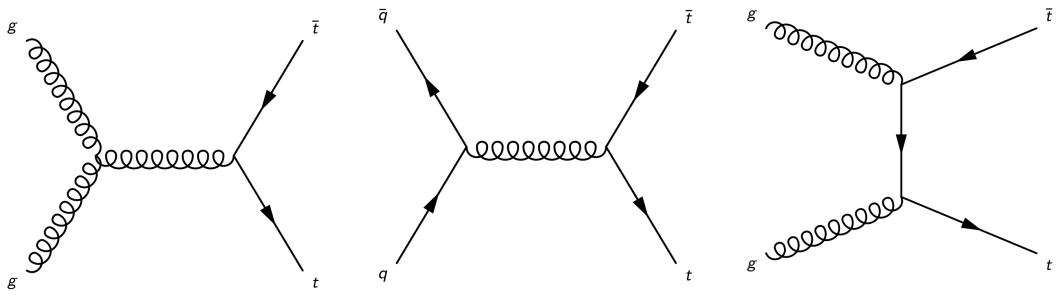
Figure 7.10: Feynman diagrams representing the WW production mechanisms available at the LHC. The two possible quark-annihilation production mechanisms are shown in the left and in the middle, while the gluon-gluon fusion is shown on the right.

which the cross-section is 12.178 pb [126], while for gg→WW channel it is 0.5905 pb [126]. In both cases, the fully leptonic decay of the two W bosons is already included in the cross-section calculation. The WW production alone represents 40% of the total amount of background of this analysis.

7.3.3 Top Production

The production of top quarks is the second largest background for this analysis, just after the WW production. Top quarks almost always decay to a b-quark and a W boson, so that the presence of b-jets is a sign of the production of top quarks in the event. In case the b-jets in the final states are not reconstructed or go to a region of the detector

out of acceptance, or are not correctly identified as produced by b-quarks, the event can mimic the signal signature if the corresponding W boson decays leptonically. The largest contribution is given by the fully leptonic decay of a pair of top quarks ($t\bar{t}$). Several production mechanisms are available for this process, as shown in Figure 7.11. The cross-



(a) (b) (c)

Figure 7.11: Feynman diagrams representing three examples of $t\bar{t}$ production mechanisms available at the LHC.

section of the $t\bar{t}$ production, considering that both W bosons coming from the decay of the top quarks decay leptonically, is 87.310 pb [126]. Another source of contamination is represented by the production of a single top, which can happen in association with a W boson (tW), or through the so-called single-top production in the s-channel or in the t-channel. The three mechanisms are illustrated in Figure 7.12. In this case, the cross-sections are smaller, and the contribution is dominated by the tW process, for which it is 7.47 pb [126]. About 40% of the total background is given by top.

7.3.4 Drell-Yan Production

The production of a Z boson, or of a virtual γ^*, at a hadron collider, leading to a final state with two same-flavour leptons, is known as Drell-Yan process. In this analysis, this process represents a source of background when the Z, or the γ^*, decay to a pair of τ leptons, which in turn decay to an electron and a muon and four neutrinos (e.g. $\tau^+ \rightarrow e^+\nu_e\bar{\nu_\tau}$ and $\tau^- \rightarrow \mu^-\bar{\nu_\mu}\nu_\tau$). The Drell-Yan cross-section is huge, but due to the small branching ratio of the decay of a τ to an electron or a muon, with respect to a quark, the contamination of this process is sensitively reduced, such that this process contributes to only 3% of the total amount of background. In Figure 7.13, the Feynman diagrams of the $Z/\gamma^* \rightarrow \tau^+\tau^-$ process and of the leptonic decay of a τ^- are shown.

7.3.5 Non-Prompt Leptons

Leptons produced inside a jet during the showering of a quark, or produced by the decay of b-quarks, can be wrongly identified as prompt leptons, namely leptons produced by the hard scattering, directly at the primary vertex of interaction. They are known as non-prompt leptons and can let enter in the signal phase space events coming from processes in which only one prompt lepton is produced. The most important of these processes is the production of a W boson in association with jets (W+jets), which is schematically shown in Figure 7.14. If the W decays leptonically and one of the jets is identified as a lepton, the final state is the same as the signal. The W+jets cross-section is much larger than the typical cross-sections involved in this analysis and corresponds to 61.5 nb [126]. On the

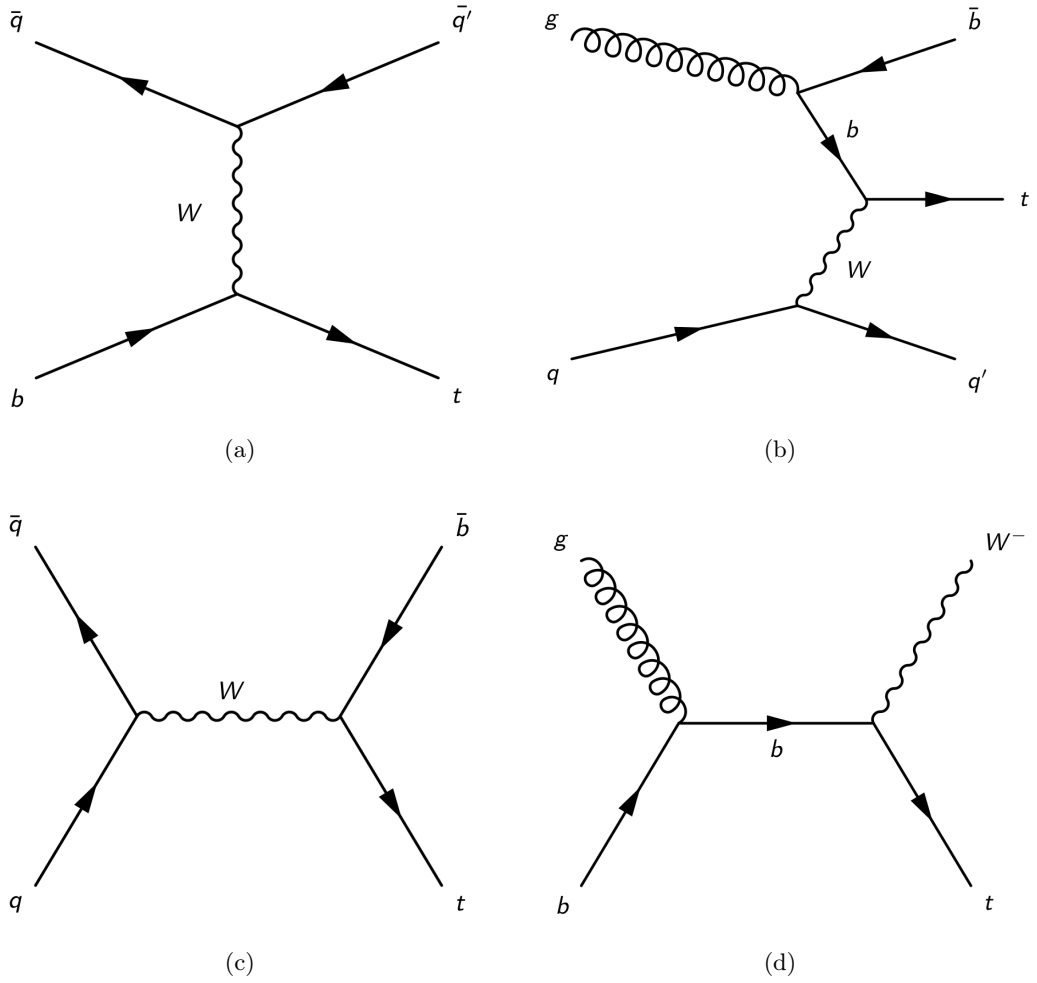


Figure 7.12: Feynman diagrams representing four examples of single-top production. On the top: two diagrams of the single-top production in the t-channel. On the bottom: on the left the single-top production in the s-channel and on the right the tW production.

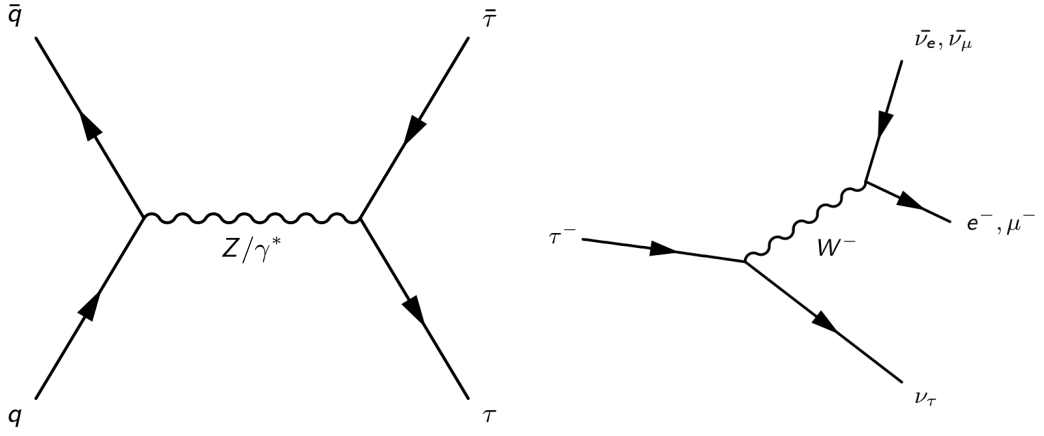
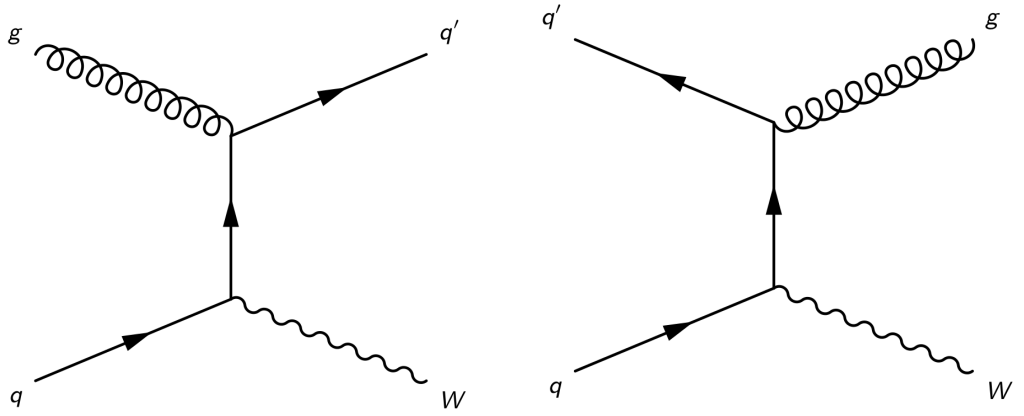


Figure 7.13: Feynman diagrams representing the $Z/\gamma^* \to \tau^+\tau^-$ process (left) and leptonic decay of a τ^- (right).

other hand, the probability to misidentify a lepton inside a jet as a prompt lepton is very low at CMS, and the level of contamination of non-prompt leptons is about 6%. Apart from W+jets, also the semi-leptonic decays of top quark pairs or of tW and single-top processes contribute to this source of background.

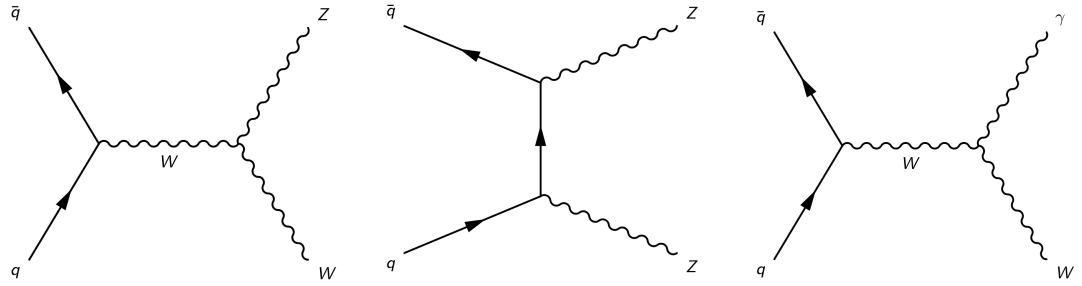


(a) (b)

Figure 7.14: Two possible mechanisms for the production of a W boson in association with a jet.

7.3.6 Di-Boson and Tri-Boson Production

Events with two or three bosons, as WZ, ZZ, WWW, WWZ, WZZ, ZZZ, can mimic the signal if they present two leptons in the final state. This is possible if the remaining leptons are not reconstructed, or in case one or more bosons decay to jets or neutrinos. Other processes that enter this category are Wγ and Wγ^*. In the former case, the photon is reconstructed as an electron, while in the latter the virtual photon decays to a pair of same flavour leptons. In Figure 7.15 three examples of Feynman diagrams for di-boson production are shown. The total signal region contamination due to multi-boson production is about 4%.



(a) WZ. (b) ZZ. (c) Wγ.

Figure 7.15: Three possible Feynman diagrams for di-boson production. From left to right: WZ, ZZ, and Wγ.

Chapter 8

Event Selection and Background Estimation

Indicare con una crocetta
la qualità, la quantità desiderata

CCCP, *CCCP*

The experimental signature of the mono-Higgs(WW) analysis presented in this document consists of two oppositely charged leptons, one electron and one muon, and large E_T^{miss}. The first requirement to select this kind of events is to choose the proper trigger paths, in this case, unprescaled di-leptonic triggers with a low p_T threshold. The lepton definition can be optimized applying identification and isolation requirements on top of the simple reconstruction criteria, reducing the background contamination. By comparing MC distributions for signal and backgrounds, additional selections can be added to enhance the signal significance, and define what is called a *signal region*, by cutting on kinematic variables that behave differently for the signal and for the main backgrounds. Once the signal region is defined, the contamination due to the background processes has to be estimated and makes use of data-driven methods whenever possible. The estimation of non-prompt leptons contamination is fully data-driven, while WW, Top, Drell-Yan, WZ and Wγ^* processes are simulated using MC, and then their normalization is computed from data in specific *control regions*. WW, Top, and Drell-Yan control regions are chosen to be orthogonal, but as close as possible to the signal region, while WZ and Wγ^* require a different approach since these processes involve three leptons in the final state. All other backgrounds rely on MC simulation, and the normalization is directly taken from the theoretical predictions. In this chapter, the trigger paths used in the analysis, the objects selections, the signal region definition, and the methods used for background estimation are discussed.

8.1 Trigger Selection

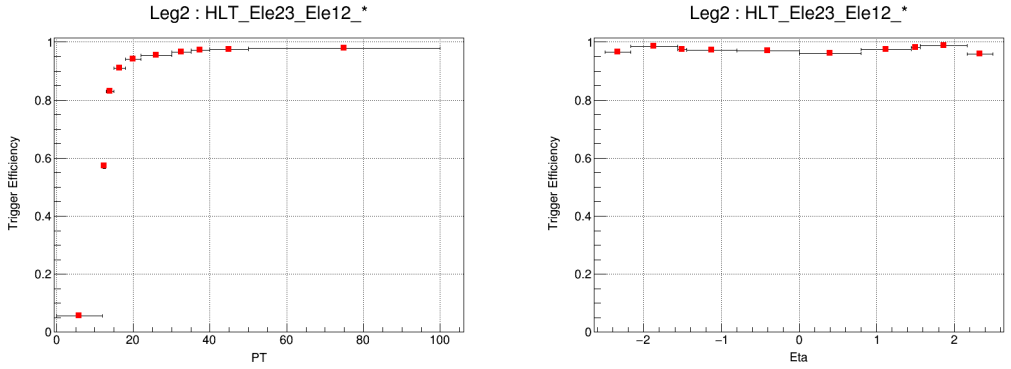
The trigger paths used for this analysis are listed in Table 8.1. The MuonEG paths select one electron and one muon with specific p_T thresholds, and have a high efficiency in collecting signal events, due to the low thresholds and the loose identification and isolation requirements. SingleElectron and SingleMuon paths present tighter requirements and are used to recover efficiency in case one lepton passes tight identification criteria, while the

second is not selected by the di-lepton path. DoubleElectron and DoubleMuon triggers are used to estimate the contamination of the WZ and Wγ^* processes in the signal region. Due to changes in data-taking conditions, the MuonEG path was modified along 2016. The higher instantaneous luminosity required the addition of a selection on the longitudinal impact parameter (d_z) for electrons, in order to avoid an increase in the trigger rate that would have made it prescaled. The same happened for the DoubleMuon paths. The

Dataset	Run range	HLT path
MuonEG	(273158,278272)	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v*
	(278273,284044)	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*
	(273158,278272)	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*
	(278273,284044)	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*
SingleElectron	(273158,284044)	HLT_Ele25_eta2p1_WPTight_Gsf_v*
	(273158,284044)	HLT_Ele27_WPTight_Gsf_v*
SingleMuon	(273158,284044)	HLT_IsoTkMu24_v*
	(273158,284044)	HLT_IsoMu24_v*
DoubleMuon	(273158,281612)	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v*
	(281613,284044)	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
	(273158,281612)	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v*
	(281613,284044)	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*
DoubleEG	(273158,284044)	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*

Table 8.1: Summary table of triggers used for the signal region.

di-lepton trigger efficiencies are measured per leg using the Tag and Probe method (see Section 6.2.1) and then combined together. Efficiencies are provided in bins of p_T and η, and are used to reweight simulated events. Examples of trigger efficiencies are shown in Figure 8.1 for electrons and Figure 8.2 for muons. The overall trigger efficiency has been measured to be about 99%. Additional triggers have been used in the analysis to select



(a) Eff vs p_T.

(b) Eff vs η.

Figure 8.1: Trigger efficiencies for the leg Ele12_* of the electron trigger path. The efficiencies are shown on the left as a function of p_T and on the right as a function of η.

a region enriched in non-prompt leptons, to measure the probability of a lepton produced inside a jet, or coming from the decay of a b-quark, to be identified as a prompt lepton (*fake rate*). These trigger paths select events with lower p_T thresholds, so that larger acquisition rates are expected, and are prescaled. They are listed in Table 8.2.

p_T/η	-2.4:-2.1	-2.1:-1.6	-1.6:-1.2	-1.2:-0.8	-0.8:-0.3	-0.3:-0.2	-0.2:0.2	0.2:0.3	0.3:0.8	0.8:1.2	1.2:1.6	1.6:2.1	2.1:2.4	max rel. st. error
10:21	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21:22	0.035	0.021	0.011	0	0	0	0	0	0	0	0.012	0.022	0.031	0.11
22:23	0.12	0.11	0.062	0.064	0.064	0.054	0.058	0.054	0.059	0.065	0.067	0.1	0.12	0.044
23:24	0.49	0.63	0.71	0.76	0.84	0.68	0.81	0.73	0.84	0.77	0.71	0.65	0.47	0.012
24:25	0.62	0.75	0.8	0.85	0.92	0.76	0.9	0.79	0.92	0.85	0.81	0.77	0.62	0.009
25:26	0.65	0.78	0.83	0.86	0.92	0.77	0.9	0.8	0.93	0.85	0.83	0.79	0.63	0.008
26:30	0.69	0.8	0.85	0.87	0.93	0.78	0.91	0.82	0.93	0.87	0.85	0.82	0.68	0.003
30:40	0.75	0.82	0.87	0.88	0.93	0.79	0.91	0.82	0.93	0.88	0.86	0.83	0.73	0.001
40:60	0.77	0.83	0.87	0.89	0.93	0.79	0.91	0.82	0.93	0.89	0.87	0.84	0.76	0.001
60:100	0.79	0.83	0.84	0.89	0.93	0.78	0.92	0.81	0.93	0.89	0.86	0.84	0.78	0.004
100:200	0.8	0.83	0.79	0.88	0.92	0.78	0.91	0.81	0.93	0.88	0.81	0.83	0.77	0.022
Color Definition														
Efficiency			0.95-1.00		0.90-0.95		0.80-0.90		0.70-0.80		0.50-0.70		< 0.50	
max rel. st. error			0.00-0.01		0.01-0.03		0.03-0.05		0.05-0.07		0.07-0.10		> 0.10	

(a) Eff vs p_T.

p_T/η	-2.4:-2.1	-2.1:-1.6	-1.6:-1.2	-1.2:-0.8	-0.8:-0.3	-0.3:-0.2	-0.2:0.2	0.2:0.3	0.3:0.8	0.8:1.2	1.2:1.6	1.6:2.1	2.1:2.4	max rel. st. error
10:13	0.91	0.92	0.96	0.94	0.94	0.85	0.93	0.81	0.94	0.93	0.96	0.92	0.93	0.005
13:16	0.92	0.92	0.97	0.95	0.94	0.83	0.94	0.84	0.94	0.95	0.96	0.93	0.94	0.004
16:20	0.93	0.93	0.98	0.95	0.96	0.85	0.95	0.86	0.95	0.95	0.97	0.93	0.94	0.002
20:25	0.93	0.93	0.97	0.96	0.96	0.86	0.95	0.87	0.96	0.95	0.97	0.94	0.95	0.001
25:30	0.94	0.93	0.98	0.96	0.96	0.87	0.96	0.87	0.96	0.95	0.97	0.94	0.95	0.001
30:40	0.94	0.93	0.98	0.95	0.96	0.87	0.95	0.87	0.96	0.95	0.97	0.93	0.95	0.0005
40:60	0.94	0.93	0.97	0.95	0.96	0.87	0.95	0.87	0.96	0.95	0.97	0.93	0.95	0.0005
60:100	0.95	0.93	0.95	0.95	0.96	0.86	0.95	0.87	0.96	0.95	0.96	0.93	0.95	0.002
100:200	0.95	0.93	0.93	0.94	0.95	0.83	0.95	0.85	0.95	0.94	0.93	0.93	0.96	0.01
Color Definition														
Efficiency			0.95-1.00		0.90-0.95		0.80-0.90		0.70-0.80		0.50-0.70		< 0.50	
max rel. st. error			0.00-0.01		0.01-0.03		0.03-0.05		0.05-0.07		0.07-0.10		> 0.10	

(b) Eff vs η.

Figure 8.2: Trigger efficiencies for the leg Mu23_* of the muon trigger path (top) and for the leg Mu8_* (bottom). The efficiencies are shown as a function of p_T and η.

Dataset	Trigger name
SingleMuon	HLT_Mu8_TrkIsoVVL_v*
	HLT_Mu17_TrkIsoVVL_v*
SingleElectron	HLT_Ele12_CaloIdM_TrackIdM_PFJet30_v*
	HLT_Ele23_CaloIdM_TrackIdM_PFJet30_v*
	HLT_Ele12_CaloIdL_TrackIdL_IsoVL_PFJet30_v*
	HLT_Ele23_CaloIdL_TrackIdL_IsoVL_PFJet30_v*

Table 8.2: Triggers used for non-prompt leptons studies.

8.2 Primary Vertex Selection

Since several proton-proton interactions happen at each bunch crossing, it is necessary to reconstruct the vertex that produced the hard scattering, defined as the primary vertex of interaction, and to remove the secondary ones. Vertices are reconstructed using the Deterministic Annealing clustering of tracks [127]. Reconstructed primary vertices are required to have a z position within 24 cm with respect to the nominal detector centre and a radial position within 2 cm with respect to the beam spot. In addition, there must be more than four degrees of freedom in the fitted vertex. From the set of primary vertices in the event passing this selection, the vertex with the largest summed squared-p_T of the associated tracks is chosen as the event primary vertex. Reconstructed leptons are required to have small impact parameters with respect to this vertex.

8.3 Muon Selection

The muon selection follows the recommendations of the CMS Muon Physics Object group (Muon POG) [128]. Muons are required to have a p_T larger than 20 GeV and to be reconstructed both in the tracker and in the muon chambers as global muons, as described in Section 5.2. This means that only muons with $|\eta| < 2.4$ are selected. In addition, several requirements ensure the quality of the muons, according to the recommended *Tight Working Point* (WP):

- **Track quality:** By selecting only tracks reconstructed with strict quality requirements, muons from decay-in-flight and punch-through (mesons which manage to reach the first muon stations and then decay) are suppressed, and reliable estimation of the p_T is ensured:
  - χ^2/**ndof** of the global-muon track fit less than 10
  - At least one **muon-chamber hit** included in the global-muon track fit
  - **Muon segments** in at least two muon stations
  - At least one **pixel hit**
  - At least five **tracker layers** with hits
- **Impact parameter:** Requiring the muon tracker track to have a small impact parameter with respect to the primary vertex helps to suppress cosmic muons and to further suppress muons from decay-in-flight. Tighter selections with respect to the ones recommended for the Tight WP are applied, in order to suppress muons from the decay of b or c hadrons:
  - **Transverse** impact parameter $\mathbf{d_{xy}} < 0.02$ cm
  - **Longitudinal** impact parameter $\mathbf{d_z} < 0.1$ cm
- **Isolation:** Two isolation selections are required:
  - The first one is set on the **particle-flow-based** isolation and takes into account corrections for the pile-up in a cone size of $\Delta R < 0.4$, defined as:

$$\mathrm{ISO} = \frac{\sum p_T^{\text{ch. had. (PV)}} + \max\left(0, \sum E_T{}^{\text{neut. had.}} + \sum E_T{}^{\text{ph.}} - 0.5 \times \sum p_T^{\text{ch. had. (PU)}}\right)}{p_T^{\mu}} \tag{8.1}$$

This selection helps in reducing the contamination from the muons originating from the hadronization of b or c quarks. For muons used in the analysis, the selection cut is $\mathrm{ISO}_{\mathrm{analysis}} < 0.15$. For the study of muons produced inside jets (non-prompt muons), a looser selection is applied: $\mathrm{ISO}_{\mathrm{non-prompt}} < 0.40$.

– The second selection is applied to the **tracker relative isolation**, which is calculated in a cone size of $\Delta\mathrm{R} < 0.3$ and defined as:

$$\mathrm{ISO}_{\mathrm{trk}} = \frac{\sum \mathrm{p_T^{tacker\ tracks}}}{\mathrm{p_T^{\mu}}} \tag{8.2}$$

This selection is used to ensure that the offline isolation ID selections are tighter than the selections applied at trigger level. Both for muons used in the analysis and for muons used for non-prompt studies, $\mathrm{ISO}_{\mathrm{trk}} < 0.4$.

In Figure 8.3 muon efficiencies in data are shown in the upper part of the canvas, while the bottom part shows Monte Carlo efficiencies. The efficiencies agree well, and from the ratio between them, the Data/MC scale factors are computed.

p_T/η	-2.4:-2.1	-2.1:-1.6	-1.6:-1.2	-1.2:-0.8	-0.8:-0.3	-0.3:-0.2	-0.2:0.2	0.2:0.3	0.3:0.8	0.8:1.2	1.2:1.6	1.6:2.1	2.1:2.4	max rel. st. error
10:13	0.87	0.93	0.93	0.87	0.95	0.84	0.94	0.82	0.93	0.87	0.92	0.95	0.9	0.01
13:16	0.9	0.95	0.94	0.9	0.95	0.86	0.96	0.84	0.95	0.9	0.94	0.95	0.89	0.005
16:20	0.9	0.95	0.96	0.91	0.96	0.85	0.95	0.83	0.96	0.92	0.96	0.96	0.91	0.003
20:25	0.91	0.96	0.97	0.92	0.97	0.85	0.96	0.84	0.96	0.92	0.97	0.96	0.92	0.002
25:30	0.92	0.96	0.97	0.93	0.97	0.85	0.96	0.84	0.97	0.93	0.97	0.97	0.92	0.001
30:40	0.92	0.96	0.97	0.93	0.97	0.86	0.96	0.84	0.97	0.93	0.97	0.97	0.92	0.0007
40:60	0.92	0.96	0.97	0.93	0.97	0.86	0.96	0.84	0.97	0.93	0.97	0.96	0.93	0.0006
60:100	0.92	0.96	0.98	0.94	0.98	0.86	0.97	0.85	0.97	0.94	0.98	0.97	0.93	0.002
100:200	0.91	0.96	0.98	0.94	0.98	0.85	0.97	0.83	0.97	0.93	0.98	0.96	0.91	0.01
Color Definition														
Efficiency			0.95-1.00		0.90-0.95		0.80-0.90		0.70-0.80		0.50-0.70		< 0.50	
max rel. st. error			0.00-0.01		0.01-0.03		0.03-0.05		0.05-0.07		0.07-0.10		> 0.10	

(a) Efficiency in data.

p_T/η	-2.4:-2.1	-2.1:-1.6	-1.6:-1.2	-1.2:-0.8	-0.8:-0.3	-0.3:-0.2	-0.2:0.2	0.2:0.3	0.3:0.8	0.8:1.2	1.2:1.6	1.6:2.1	2.1:2.4	max rel. st. error
10:13	0.93	0.96	0.96	0.95	0.96	0.84	0.96	0.9	0.98	0.96	0.97	0.96	0.92	0.011
13:16	0.93	0.97	0.98	0.95	0.98	0.88	0.96	0.86	0.97	0.96	0.98	0.97	0.94	0.008
16:20	0.95	0.98	0.99	0.96	0.98	0.9	0.97	0.86	0.98	0.96	0.99	0.98	0.95	0.005
20:25	0.95	0.99	0.99	0.97	0.99	0.9	0.97	0.89	0.98	0.97	0.99	0.99	0.95	0.003
25:30	0.95	0.99	0.99	0.97	0.99	0.9	0.98	0.89	0.99	0.97	0.99	0.99	0.96	0.002
30:40	0.95	0.99	0.99	0.97	0.99	0.9	0.98	0.89	0.99	0.97	0.99	0.99	0.96	0.001
40:60	0.96	0.98	0.99	0.97	0.99	0.9	0.98	0.89	0.99	0.97	0.99	0.99	0.96	0.0009
60:100	0.96	0.99	0.99	0.97	0.99	0.91	0.98	0.89	0.99	0.97	0.99	0.98	0.97	0.003
100:200	0.93	0.98	1	0.96	0.99	0.88	0.98	0.89	0.99	0.97	0.99	0.98	0.95	0.017
Color Definition														
Efficiency			0.95-1.00		0.90-0.95		0.80-0.90		0.70-0.80		0.50-0.70		< 0.50	
max rel. st. error			0.00-0.01		0.01-0.03		0.03-0.05		0.05-0.07		0.07-0.10		> 0.10	

(b) Efficiency in MC.

Figure 8.3: Efficiencies for Muons Tight WP in data (top) and Monte Carlo (bottom). Efficiencies are shown in bins of $\mathrm{p_T}$ and η.

8.4 Electron Selection

The electron selection follows the recommendations of the CMS EGamma Physics Object Group (EGM POG) [129]. Electrons selected in this analysis are required to have a $\mathrm{p_T}$ larger than 20 GeV. Several variables that provide discriminating power are used to identify electrons, and are grouped into three main categories:

- **Calorimetric observables:** Used to separate genuine electrons (signal electrons or electrons from photon conversions) from misidentified electrons (jets with large electromagnetic components), based on the transverse shape of electromagnetic showers in the ECAL and exploiting the fact that electromagnetic showers are narrower than hadronic showers. The energy fractions deposited in the HCAL (expected to be small, as electromagnetic showers are essentially fully contained in the ECAL), as well as the energy deposited in the preshower in the endcaps are also used.

  - **hOverE ($\frac{\mathrm{H}}{\mathrm{E}}$):** Ratio of energy deposited in HCAL to ECAL
  - **ooEmoop ($\frac{1}{\mathrm{E_{SC}}} - \frac{1}{\mathrm{p}}$):** $\mathrm{E_{SC}}$ is the Super Cluster (SC) energy and p is the track momentum at the point of closest approach to the vertex
  - **sigmaIetaIeta ($\sigma_{\eta\eta}$):** The weighted cluster RMS along η inside 5×5 region of supercluster
  - **dEtaIn ($\Delta\eta$):** η difference between supercluster and position of inner track extrapolated from interaction vertex
  - **dPhiIn ($\Delta\phi$):** ϕ difference between supercluster and position of inner track extrapolated from interaction vertex

- **Isolation variables:** A significant fraction of background to isolated primary electrons is due to misidentified jets or to genuine electrons within a jet resulting from semileptonic decays of b or c quarks. In both cases, the electron candidates have significant energy flow near their trajectories, and requiring electrons to be isolated from such nearby activity greatly reduces these sources of background.

  - **relIsoWithEA:** PF based isolation defined in a cone size of 0.3 in the (η,ϕ) plane around the electron direction, relative to electron $\mathrm{p_T}$. The contribution to the isolation due to pile-up is subtracted by using the *effective area* (EA) method, by removing to the isolation computation the expected amount of energy deposit in the cone area due to secondary vertices:

    $$\mathrm{ISO_{EA}} = \frac{\sum \mathrm{p_T^{ch.had.(PV)}} + \max(0, \sum \mathrm{E_T^{neut.had.}} + \sum \mathrm{E_T^{ph.}} - \rho \times (\mathrm{EA^{ph.}} + \mathrm{EA^{neut.had.}}))}{\mathrm{p_T^e}} \tag{8.3}$$

    Where ρ represents the energy density associated with the pile-up and EA is the effective area of the isolation cone, which depends on the number of vertices in the event, on η, and on the pile-up component (photons, or neutral hadrons).

- **Conversion rejection variables:** Used to reject secondary electrons produced from photon conversions. To reject this background, CMS algorithms exploit the pattern of track hits. When photon conversions take place inside the volume of the tracker, the first hit on electron tracks from the converted photons is often not located in the innermost layer of the tracker, and missing hits are therefore present in that region. For prompt electrons, whose trajectories start from the beamline, no missing hits are expected in the inner layers. Also, the impact parameter of the electron is required to be small for the electron to be originated from the vertex of interest.

  - **Conversion veto**
  - **Number of missing inner hits**
  - **Transverse impact parameter ($\mathrm{d_{xy}}$)**

– **Longitudinal impact parameter (d_z)**

A set of selections is applied to the variables defined above to separate the real electrons from fakes. A set of cuts on those variables for achieving a particular signal efficiency are applied accordingly to the recommended Tight WP. The cuts are different for electrons in the ECAL Barrel (EB) and in the Endcap (EE), and electrons are required to have $|\eta| < 2.5$. On top of that, a set of cuts to make the electron ID selection *Trigger safe* is applied, namely to ensure that ID selections are tighter than trigger ones. Table 8.3 shows the official Tight WP of Electron ID and Table 8.4 shows the extra cuts applied. These extra cuts are also known as *WP Loose* in names of HLT triggers. Electron triggers have some identification and isolation requirements at HLT level and the isolation used here is detector-based. The offline working point of electron identification uses particle flow based isolation. It is not trivial to compare the efficiencies of these two isolation approaches. Hence it is safe to use these cuts in the basic identification of electrons in the analysis. Along with these selections, the missing hits cut has been tightened to suppress mainly the Wγ background. In Figure 8.4 data efficiencies are shown in the upper part of the

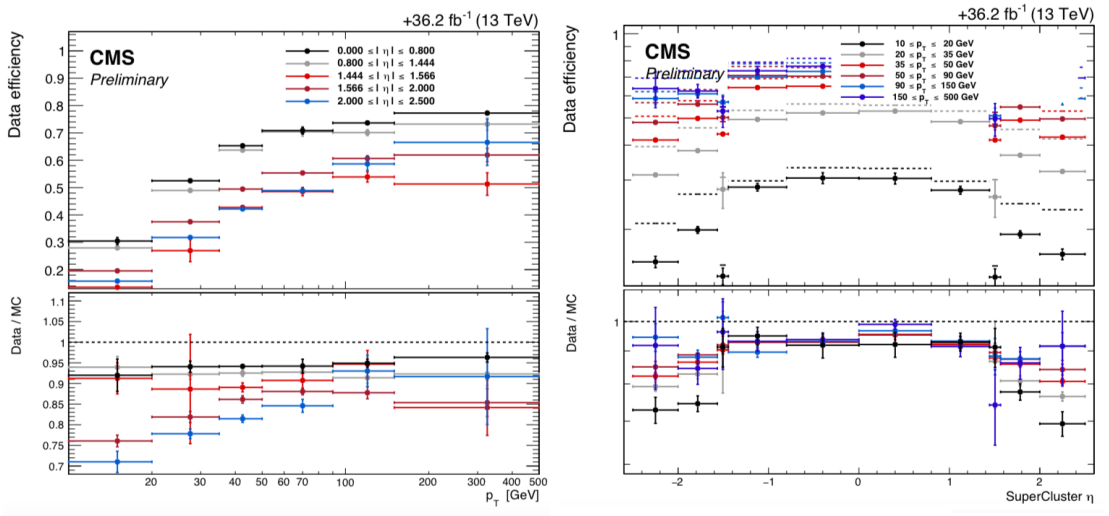
Id Variable		Cut for EB	Cut for EE
Full 5×5 $\sigma_{\eta\eta}$	$<$	0.00998	0.0292
$\lvert\Delta\eta_{\mathrm{Seed-Calo}}\rvert$	$<$	0.00308	0.00605
$\lvert\Delta\phi_{\mathrm{Seed-Calo}}\rvert$	$<$	0.0816	0.0394
$\frac{\mathrm{H}}{\mathrm{E}}$	$<$	0.0414	0.0641
Rel Iso With EA	$<$	0.0588	0.0571
$\left\lvert\frac{1}{\mathrm{E_{SC}}} - \frac{1}{\mathrm{p}}\right\rvert$	$<$	0.0129	0.0129
$\lvert\mathrm{d_{xy}}\rvert$	$<$	0.05	0.1
$\lvert\mathrm{d_z}\rvert$	$<$	0.1	0.2
Missing Inner Hits	$\leq$	1	1
Pass Conversion Veto		yes	yes

Table 8.3: Cut-Based Tight electron ID working point for Barrel ($|\eta_{\mathrm{SC}}| \leq 1.479$) and for Endcap ($1.479 < |\eta_{\mathrm{SC}}| < 2.5$).

Id Variable		Cut for EB	Cut for EE
ECAL PF Cluster Iso	$<$	0.160	0.120
HCAL PF Cluster Iso	$<$	0.120	0.120
Tracker Iso	$<$	0.08	0.08
$\frac{H}{E}$	$<$	0.060	0.065
full 5x5 $\sigma_{\eta\eta}$	$<$	0.011	0.031
$\lvert\Delta\eta_{\mathrm{Seed-Calo}}\rvert$	$<$	0.004	-
$\lvert\Delta\phi_{\mathrm{Seed-Calo}}\rvert$	$<$	0.020	-
$\left\lvert\frac{1}{\mathrm{E_{SC}}} - \frac{1}{\mathrm{p}}\right\rvert$	$<$	0.013	0.013
GsfTrackc χ^2/NDOF	$<$	-	3.0
Missing Inner Hits	$<$	1	1
$\lvert\mathrm{d_z}\rvert$	$<$	0.373	0.602
$\lvert\mathrm{d_{xy}}\rvert$	$<$	0.1	0.2
Pass Conversion Veto		yes	yes

Table 8.4: Additional cuts applied to electrons in the mono-Higgs(WW) analysis.

canvas, while the bottom part shows the scale factor as a function of p_T for various η bins and vice-versa. For the barrel part, the scale factors are very close to unity, however, for the endcap, they are 20-40% off from unity.



(a) p_T efficiencies.

(b) η efficiencies.

Figure 8.4: Efficiencies for Electrons cut based Tight WP + mono-Higgs(WW) cuts as a function of p_T (left) and η (right). Efficiencies for data are shown in the upper part of the plots and scale factors in the lower part.

8.5 Jet Selection

The jet selection follows the recommendations of the CMS Jet and Missing Transverse Energy Physics Object Group (JET/MET POG) [130]. The Loose WP definition is applied, as described in Table 8.5. In addition, jets are required to have a distance in the (η,ϕ) plane of at least 0.3 (ΔR(jet,lepton) $>$ 0.3) with respect to any reconstructed lepton, in order to avoid double counting of objects as leptons and jets. No explicit selections on the

Id Variable		Selection
Neutral hadron fraction	$<$	0.99
Neutral electromagnetic fraction	$<$	0.99
Number of constituents	$>$	1
Charged hadron fraction	$>$	0
Charged multiplicity	$>$	0
Charged electromagnetic fraction	$<$	0.99

Table 8.5: Jet Loose ID working point selection.

number of jets in an event are required. Instead, a veto on the presence of b-jets is imposed. Also in this case, the definition of b-jet is taken by the corresponding CMS Physics Object Group recommendations (B-Tagging and Vertexing, BTV [131]). The cMVAv2 b-tagging algorithm, described in Section 5.4.1, has been used, and the selection follows the Loose WP, to enhance the rejection of top-quark events. A jet is considered as b-tagged when:

- p_T ^{jet} > 20 GeV
- $|\eta^{jet}| < 2.4$
- cMVAv2 > -0.5884

8.6 E_T^{miss} Selection

Two different definitions of missing transverse energy are considered in this analysis. The first one goes under the name of Particle-flow E_T^{miss} with Type-I corrections. It is defined as the negative sum of the momenta of all the particle-flow candidates in an event and takes into account the Jet Energy Corrections, described in Section 5.4. Furthermore, following the recommendations of the CMS JET/MET POG, a set of filters is applied in order to reject events with an unphysically large amount of missing energy [132]. These filters efficiently remove events in which the E_T^{miss} is not correctly measured due to:

- Detection of beam halo in the event, taken into account in the E_T^{miss} implementation
- Noise in the calorimeters
- Bad muon or charged hadron reconstruction

To suppress events coming from processes in which there is real missing energy due to the presence of neutrinos, but the kinematic distributions are different, the projected E_T^{miss} is introduced:

$$\text{proj.}E_T^{miss} = \begin{cases} E_T^{miss}, & \text{if} \Delta\phi_{min} \geq \frac{\pi}{2} \\ E_T^{miss} \sin(\Delta\phi_{min}), & \text{if} \Delta\phi_{min} < \frac{\pi}{2} \end{cases} \tag{8.4}$$

where:

$$\Delta\phi_{min} = \min(\Delta\phi(E_T^{miss}, \ell_1), \Delta\phi(E_T^{miss}, \ell_2)) \tag{8.5}$$

The projected E_T^{miss} is particularly effective in rejecting $Z/\gamma^* \to \tau^+\tau^-$ events. In this process, the large mass of the Z boson with respect to τ leptons, makes the taus strongly boosted, so that their decay products (leptons and neutrinos) are aligned together. Since in this case, the only sources of E_T^{miss} are neutrinos, taus and E_T^{miss} will be aligned, as shown in Figure 8.5. In the signal case, the angle between leptons and E_T^{miss} is typically larger, so that a different projected E_T^{miss} distribution is expected, as illustrated in Figure 8.6. The second E_T^{miss} definition, known as Tracker E_T^{miss} uses only the momenta of particles measured by tracker, in order to reduce the dependence on the pile-up:

$$\text{tracker } E_T^{miss} = -\vec{p_T}^{\ell_1} - \vec{p_T}^{\ell_2} - \sum_j \vec{p_T}^{j} \tag{8.6}$$

Where $\vec{p_T}^{\ell_1}$ and $\vec{p_T}^{\ell_2}$ are the momenta of the two leptons with higher p_T passing the analysis selections, and the sum on $\vec{p_T}^{j}$ considers the particle-flow candidates satisfying the following conditions:

- The track matched to the PF candidate has a longitudinal impact parameter $|d_z| <$ 0.1 cm with respect to the primary vertex
- The track has a distance in the (η,ϕ) plane $\Delta R > 0.1$ to both leptons, to avoid double counting of the track

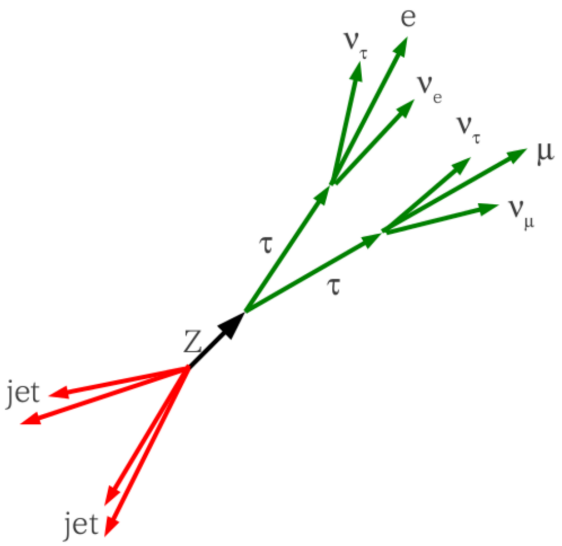
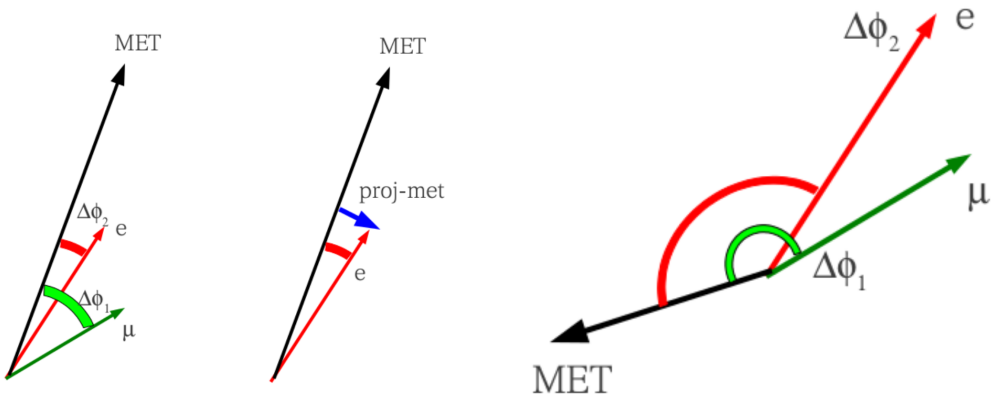


Figure 8.5: Schematic representation of the $Z/\gamma^* \to \tau^+\tau^-$ process.



(a) $Z/\gamma^* \to \tau^+\tau^-$ proj. E_T^{miss}.

(b) Signal proj. E_T^{miss}.

Figure 8.6: Expected projected E_T^{miss} in $Z/\gamma^* \to \tau^+\tau^-$ process (left) and signal (right). In the background case, the minimum of the angles between E_T^{miss} and leptons is less than $\frac{\pi}{2}$, so that the E_T^{miss} is projected on the transverse plane. In the signal case, the angle between the E_T^{miss} and the leptons is expected to be large, so that the E_T^{miss} is left untouched.

Also for the tracker E_T^{miss}, the projection on the leptons momenta is introduced. The projected PF E_T^{miss} and the projected tracker E_T^{miss} are almost fully correlated for the signal, but not for the background, so that the minimum of the two is used in the analysis to enhance the signal selection:

$$\text{min . proj. } E_T^{miss} = \text{min}(\text{proj. P.F. Type I } E_T^{miss}, \text{proj. track. } E_T^{miss}) \tag{8.7}$$

The minimum projected E_T^{miss} is also defined as *mpmet*.

8.7 Signal Region Definition

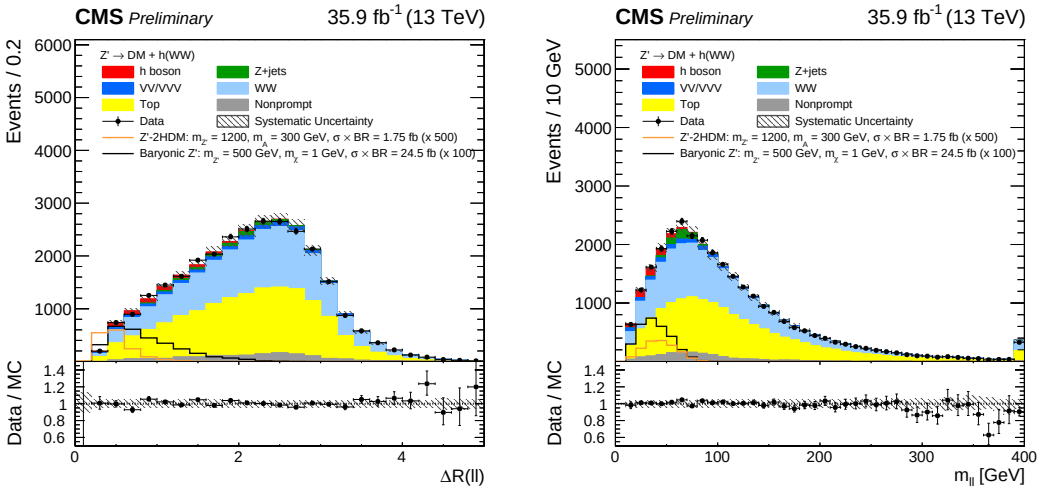
To select a phase space enriched in signal events, the basic requirements are the presence of two well-identified and isolated leptons with opposite charge and different flavour (one electron and one muon) and a large amount of missing transverse energy, due to the presence of the undetected neutrinos and DM particles. Additional selection cuts, used to enhance the signal to background ratio, are applied following the traditional approach of the SM W^+W^- [133] or $h \to W^+W^-$ [134] analyses, with the difference that in this case no selection on the number of jets in the event is required. This selection is in fact affected by large theoretical uncertainties and is used to reduce the contamination of $t\bar{t}$ events. Since this process does not represent an overwhelming source of background for this analysis, and to avoid the propagation of the related uncertainties, the jet selection is dropped and the analysis is inclusive in the number of jets. The leading lepton is required to have a p_T larger than 25 GeV, in order to have the triggers used to select events fully efficient. Due to the fact that the Higgs boson is typically highly boosted, the trailing lepton threshold is set to 20 GeV without losing signal efficiency. This selection allows reducing the contamination of non-prompt leptons, from W + jets or semi-leptonic $t\bar{t}$ decay, as leptons inside jets are often softer. To further suppress non-prompt lepton contamination, a selection on the p_T of the di-lepton system larger than 30 GeV is introduced. Events with a third loosely identified lepton with p_T larger than 10 GeV are rejected to suppress tri-boson processes. Loose leptons are defined in the same way as leptons used for non-prompt studies and described in Sections 8.3 and 8.4. A di-lepton invariant mass larger than 12 GeV is required, in order to get rid of low-mass resonances. To reduce the contamination of $Z/\gamma^* \to \tau^+\tau^-$ events in the signal region, mpmet must be larger than 20 GeV and the transverse mass of the Higgs boson ($m_T^{\ell\ell E_T^{miss}}$) must be larger than 40 GeV. This variable is defined as:

$$m_T^{\ell\ell E_T^{miss}} = \sqrt{2 \cdot p_T^{\ell\ell} \cdot E_T^{miss} \cdot \left(1 - \cos\Delta\phi(\ell\ell, E_T^{miss})\right)} \tag{8.8}$$

Finally, the top quark processes (fully leptonic $t\bar{t}$ and tW) are rejected by requiring no events with loosely identified b-jets. A jet is identified as a loose b-jet if it has a p_T larger than 20 GeV and a cMVAv2 discriminant larger than -0.5884. This set of selections defines the so-called pre-selection phase space and is summarized in the upper part of Table 8.6. Two plots of the pre-selection phase space are shown in Figure 8.7, where it can be appreciated that a good agreement between data and simulation is obtained at this level. By studying these kinematic distributions, two more selections have been applied on top of the previous ones: the di-lepton invariant mass is required to be lower than 76 GeV, as the Higgs resonance, spoiled by the presence of neutrinos, peaks at low values; the distance between the two leptons in the (η,ϕ) plane has to be lower than 2.5, due to the boost of the Higgs recoiling against the dark matter system. These two selection cuts complete the definition of the signal region, and are summarized in the two last lines of Table 8.6. Control plots for several kinematic distributions in the signal region are shown in Figures 8.8, 8.9 and 8.10. Here the transverse masses of the leading and the trailing W bosons have been introduced. For a final state with two leptons and E_T^{miss}, the transverse mass of a W boson is built by using the corresponding lepton and the full E_T^{miss}, as:

$$m_T^W = \sqrt{2 \cdot p_T^{\ell} \cdot E_T^{miss} \cdot \left(1 - \cos\Delta\phi(\ell, E_T^{miss})\right)} \tag{8.9}$$

A general good agreement between data and simulation is found. By comparing the signal and background distributions, it can be seen that each variable has still some discriminating power, but none of them can alone provide a definitive separation. This fact is exploited for the extraction of the signal, as explained in Chapter 9.

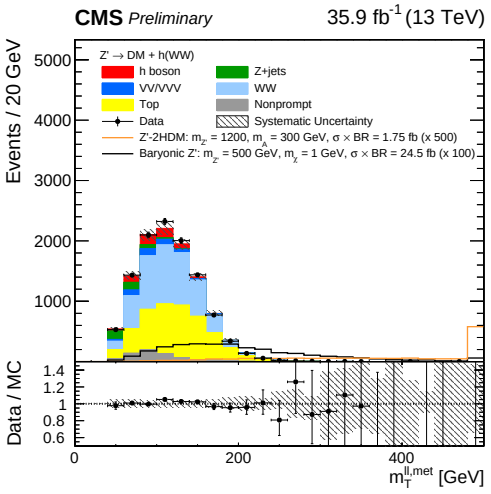


(a) $\Delta R(\ell, \ell)$. (b) $m_{\ell\ell}$.

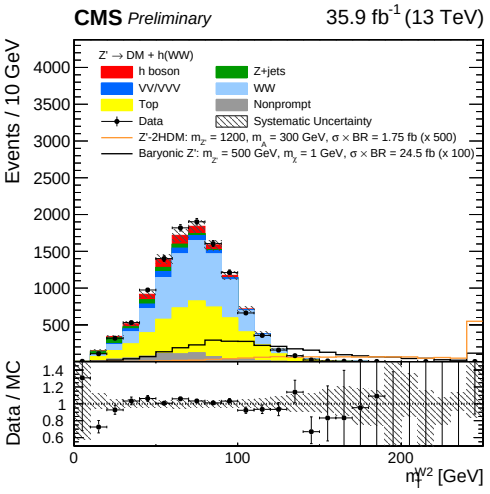
Figure 8.7: Distributions of the angular distance in the (η,ϕ) plane between the two leptons $\Delta R(\ell, \ell)$ (left) and of the di-lepton invariant mass $m_{\ell\ell}$ (right) at pre-selection level. Signal events are scaled by a factor 100 for the Z' baryonic model and by a factor 500 for the Z'-2HDM model.

Variable	Selection		Reason
$q_{\ell_1} \times q_{\ell_2}$	$<$	0	Higgs boson decays to W^+W^-
$p_T^{\ell_1}$ [GeV]	$>$	25	Couple to trigger selections
$p_T^{\ell_2}$ [GeV]	$>$	20	Suppress non-prompt leptons contamination
$p_T^{\ell_3}$ [GeV]	$<$	10	Suppress tri-boson events
$m_{\ell\ell}$ [GeV]	$>$	12	Suppress low resonances
$p_T^{\ell\ell}$ [GeV]	$>$	30	Suppress non-prompt leptons contamination
pf E_T^{miss} [GeV]	$>$	20	Suppress $Z/\gamma^* \to \tau^+\tau^-$ contamination
mpmet [GeV]	$>$	20	Suppress $Z/\gamma^* \to \tau^+\tau^-$ contamination
$m_T^{\ell\ell E_T^{miss}}$ [GeV]	$>$	40	Suppress $Z/\gamma^* \to \tau^+\tau^-$ contamination
Jets cMVAv2	$<$	-0.5884	Suppress top quark contamination
$m_{\ell\ell}$ [GeV]	$<$	76	Suppress non-resonant processes
$\Delta R(\ell, \ell)$ [rad]	$<$	2.5	Suppress non-boosted events

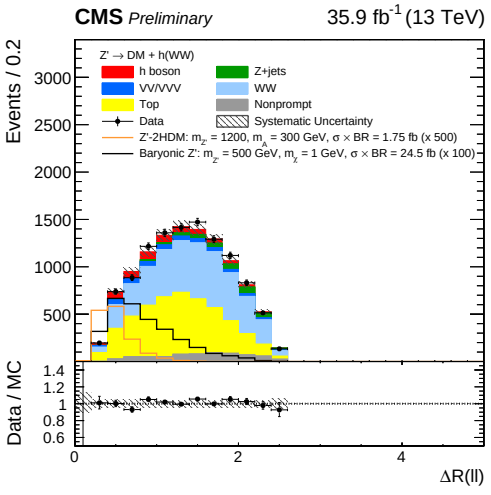
Table 8.6: Summary of the mono-Higgs(WW) signal region definition.



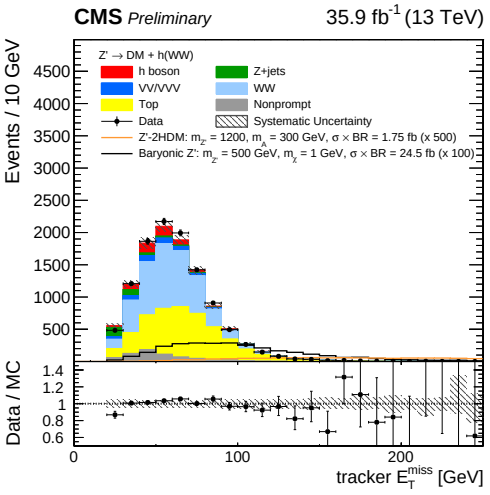
(a) $m_T^{\ell\ell E_T^{miss}}$.



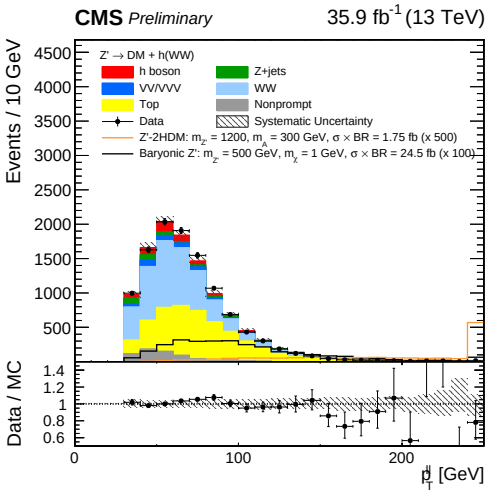
(b) $m_T^{W_2}$.



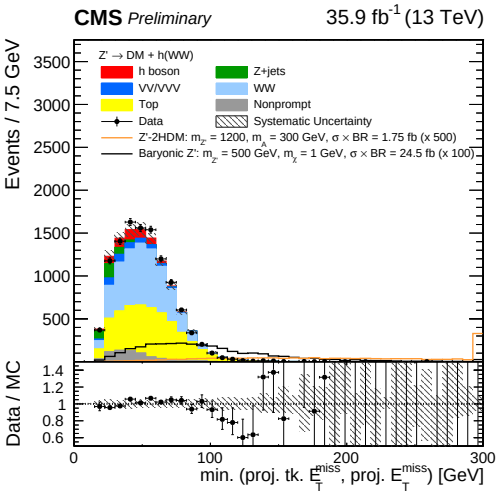
(c) $\Delta R(\ell,\ell)$.



(d) track. E_T^{miss}.

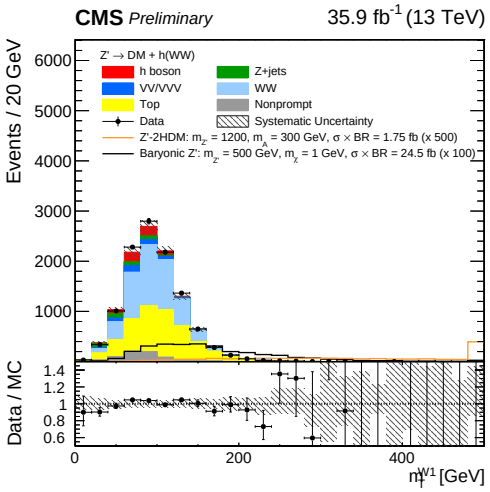


(e) $p_T^{\ell\ell}$.

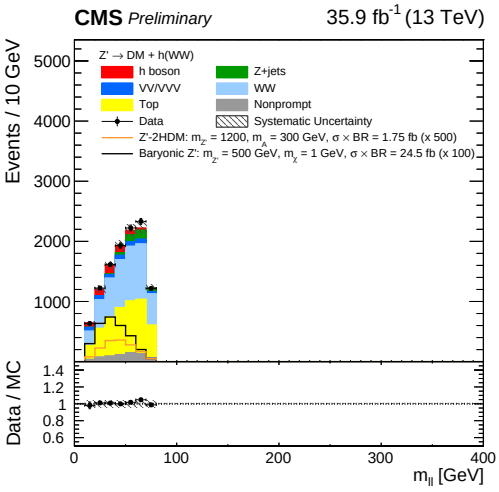


(f) mpmet.

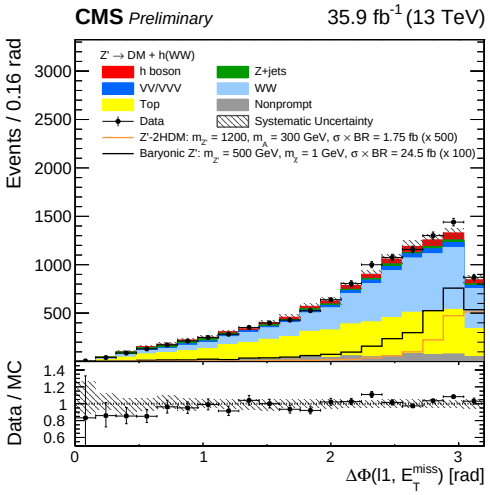
Figure 8.8: Control plots for several variables in the signal region. Good agreement between data and simulation is observed. Signal events are scaled by a factor 100 for the Z' baryonic model and by a factor 500 for the Z'-2HDM model.



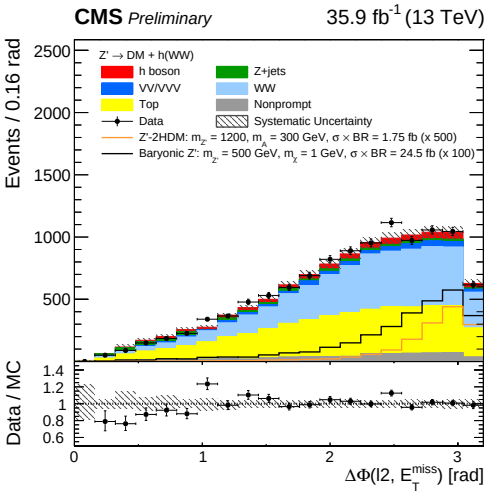
(a) $m_T^{W_1}$.



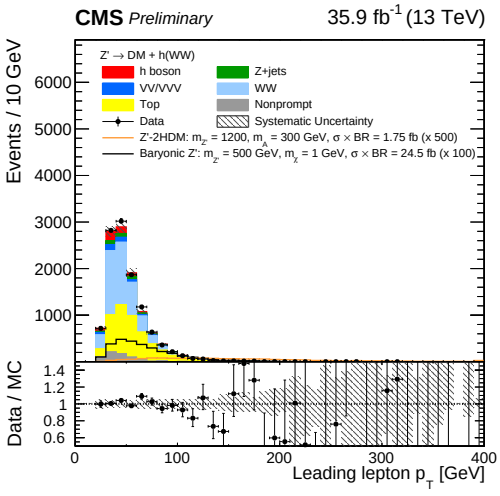
(b) $m_{\ell\ell}$.



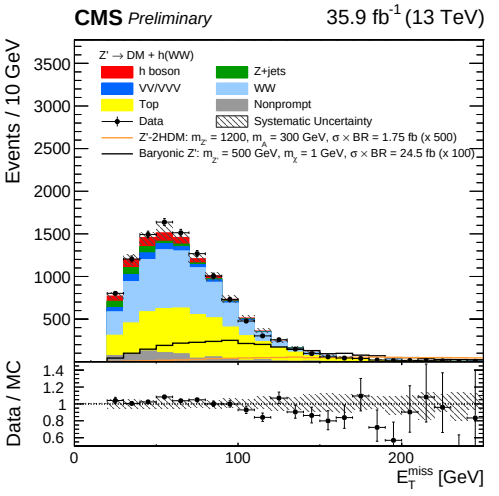
(c) $\Delta\phi(\mathrm{E_T^{miss}}, \ell_1)$.



(d) $\Delta\phi(\mathrm{E_T^{miss}}, \ell_2)$.

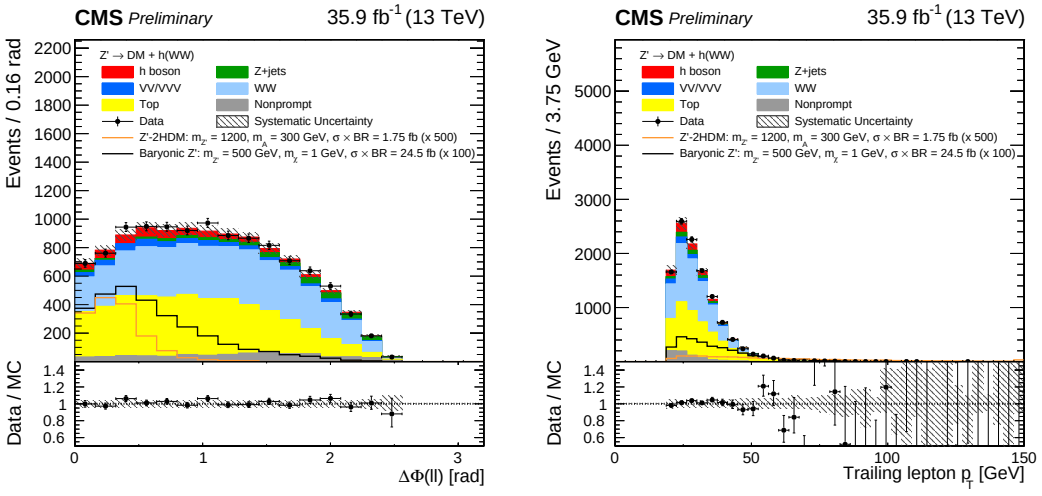


(e) leading lepton $\mathrm{p_T}$.



(f) P.F.Type I $\mathrm{E_T^{miss}}$.

Figure 8.9: Control plots for several variables in the signal region. Good agreement between data and simulation is observed. Signal events are scaled by a factor 100 for the Z' baryonic model and by a factor 500 for the Z'-2HDM model.



(a) $\Delta\phi(\ell,\ell)$.

(b) trailing lepton p_T.

Figure 8.10: Control plots for several variables in the signal region. Good agreement between data and simulation is observed. Signal events are scaled by a factor 100 for the Z' baryonic model and by a factor 500 for the Z'-2HDM model.

8.8 WW Estimation

The WW process is the main source of background for this search. The $gg \to W^+W^-$ production is simulated at LO using MCFM [135], and the cross-section is reweighted to NNLO [126]. The $qq \to W^+W^-$ production, which is much more abundant, is simulated with POWHEG at NLO. The cross-section is then corrected to NNLO [136] accuracy, and the transverse momentum of the WW system is reweighted to NNLO + NNLL accuracy [137, 138]. In Figure 8.11 the effect of the reweighting is shown. In this reweighting procedure, the efficiency uncertainties coming from higher order contributions are estimated in the different bins of the final distribution using the NNLO+NNLL variations, due to both QCD and resummation scale. Since the two W bosons in this case do not come from the decay of a Higgs, the di-lepton invariant mass is not constrained to low values. In addition, the WW pair can be weakly boosted, so that the two leptons are not expected to be as close as in the signal case. A control region enriched in WW events is thus defined by inverting the $m_{\ell\ell}$ selection of the signal region and by relaxing the $\Delta R(\ell,\ell)$ selection, in order to have a better estimation of the normalization of the process. The definition of the WW control region is summarized in Table 8.7. In Figures 8.12 and 8.13 several kinematic distributions for the WW control region are shown, and a good agreement between data and simulation is observed.

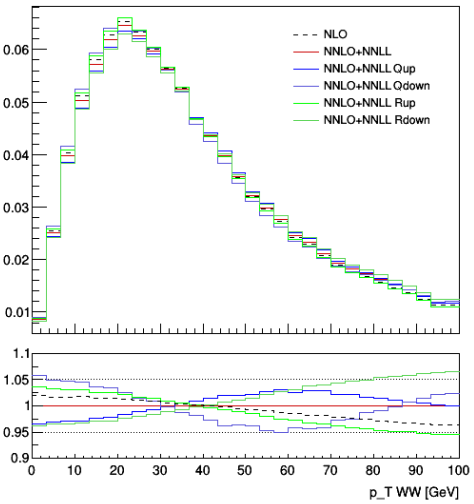
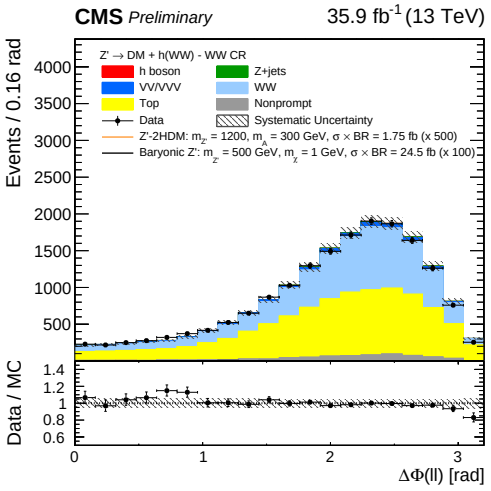


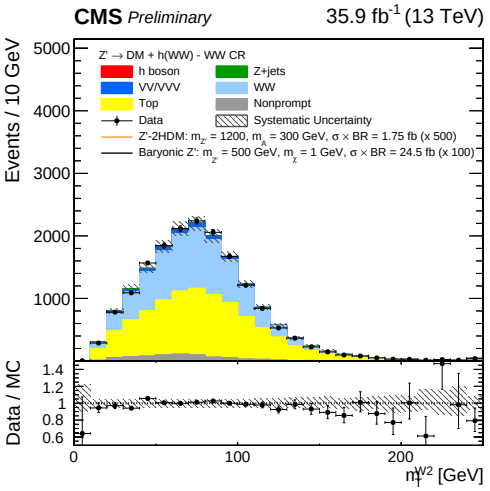
Figure 8.11: Effect of WW p_T reweight on the p_T distribution. The distribution before reweighting, the reweighted distribution, and the scale variations (R = resummation, Q = QCD scale) are shown as well.

Variable	Selection	
$q_{\ell_1} \times q_{\ell_2}$	$<$	0
$p_T^{\ell_1}$ [GeV]	$>$	25
$p_T^{\ell_2}$ [GeV]	$>$	20
$p_T^{\ell_3}$ [GeV]	$<$	10
$m_{\ell\ell}$ [GeV]	$>$	12
$p_T^{\ell\ell}$ [GeV]	$>$	30
pf E_T^{miss} [GeV]	$>$	20
mpmet [GeV]	$>$	20
$m_T^{\ell\ell E_T^{miss}}$ [GeV]	$>$	40
Jets cMVAv2	$<$	-0.5884
$m_{\ell\ell}$ [GeV]	$>$	76
$\Delta R(\ell, \ell)$		-

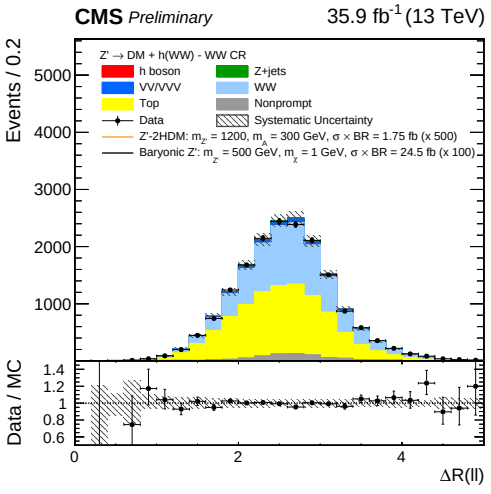
Table 8.7: Summary of the WW control region definition. The selections that change with respect to the signal region are put in the bottom of the table.



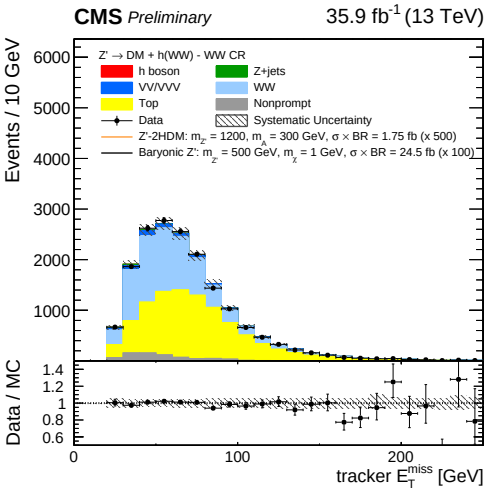
(a) $\Delta\phi(\ell,\ell)$.



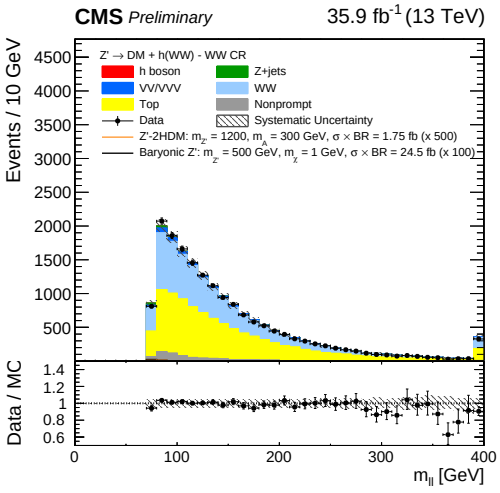
(b) $\mathrm{m_T^{W_2}}$.



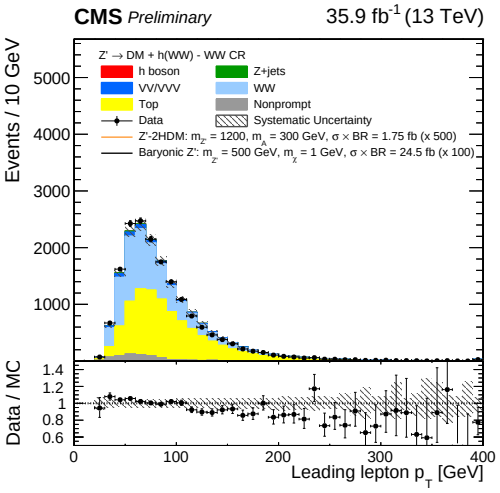
(c) $\Delta\mathrm{R}(\ell,\ell)$.



(d) track. $\mathrm{E_T^{miss}}$.

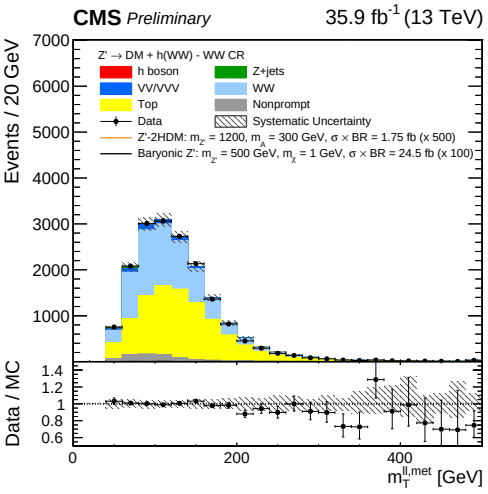


(e) $\mathrm{m}_{\ell\ell}$.

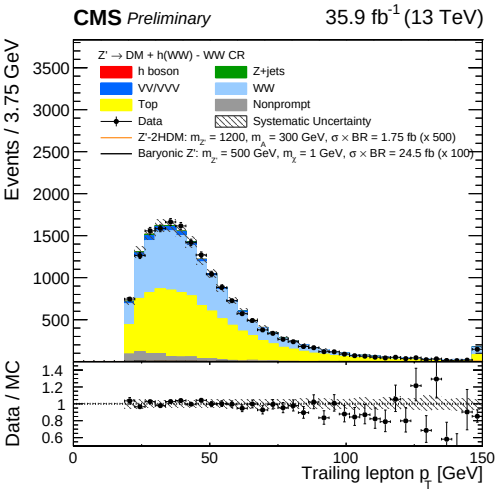


(f) leading lepton $\mathrm{p_T}$.

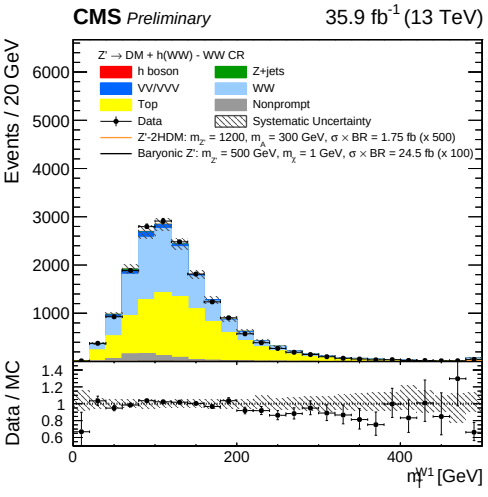
Figure 8.12: Control plots for several variables in the WW control region.



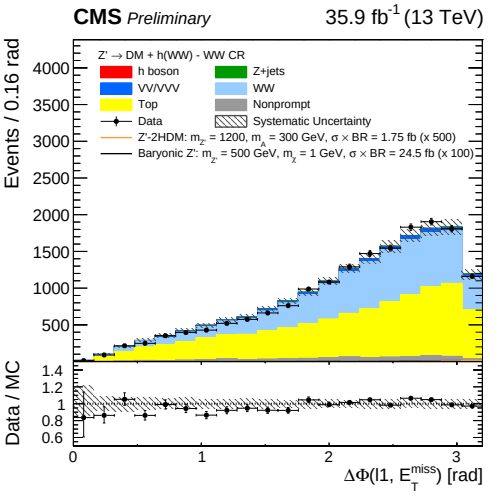
(a) $\mathrm{m_T^{\ell\ell E_T^{miss}}}$.



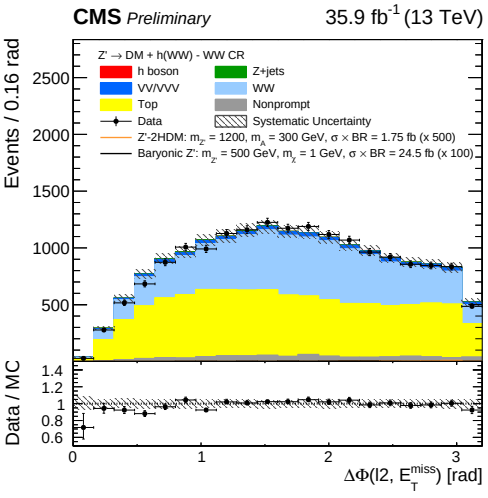
(b) trailing lepton $\mathrm{p_T}$.



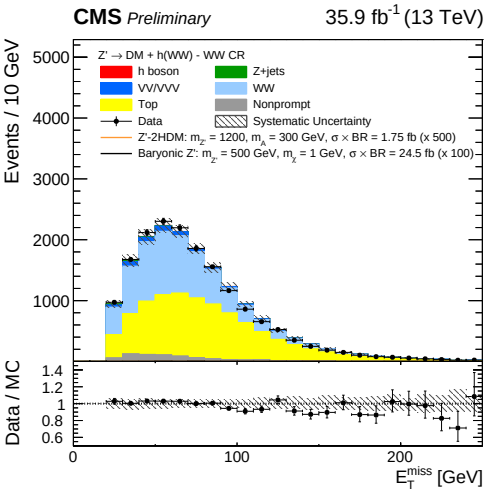
(c) $\mathrm{m_T^{W_1}}$.



(d) $\Delta\phi(\mathrm{E_T^{miss}}, \ell_1)$.



(e) $\Delta\phi(\mathrm{E_T^{miss}}, \ell_2)$.



(f) $\mathrm{E_T^{miss}}$.

Figure 8.13: Control plots for several variables in the WW control region.

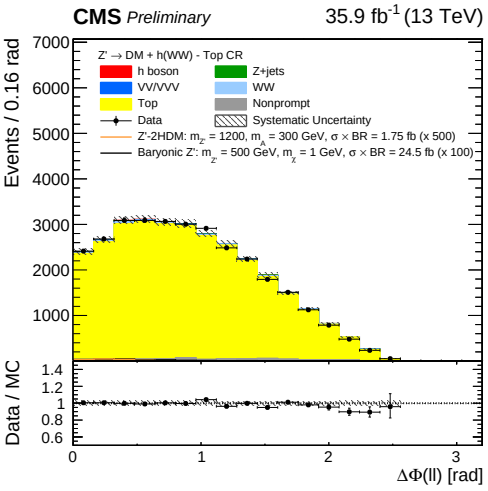
8.9 Top Estimation

The second largest background for the mono-Higgs(WW) analysis is the production of top quarks ($t\bar{t}$, tW, and single top). It is simulated by using POWHEG at NLO accuracy. Since top quarks decay almost always to a W boson and a b-quark, the peculiarity of this background is the presence of b-jets in the final state, with a relatively high transverse momentum. Therefore, b-jet identification is fundamental to identify and reject top events. The tagging efficiencies and the mistag rates (the probability of tagging a jet produced by a light quark as a b-jet) can be different in data and simulation. Scale factors are computed in order to correct these discrepancies, are applied to MC as weights per event, and depend on the number of jets, on their flavour, and on the jet kinematics (p_T, η). This procedure allows to correct for shape differences between data and MC, but does not ensure that the overall normalization is properly reproduced. A control region enriched in top event is defined, in order to adjust the normalization and to constrain uncertainties affecting the process. Such control region is defined by inverting the b-veto selection of the signal region, namely requiring that the leading jet in the event is b-tagged, as shown in Table 8.8. In

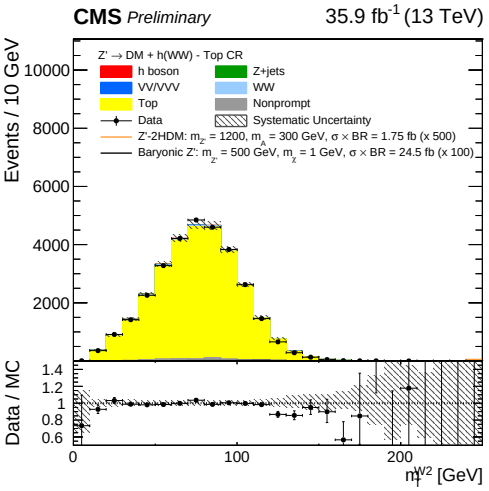
Variable	Selection	
$q_{\ell_1} \times q_{\ell_2}$	$<$	0
$p_T^{\ell_1}$ [GeV]	$>$	25
$p_T^{\ell_2}$ [GeV]	$>$	20
$p_T^{\ell_3}$ [GeV]	$<$	10
$m_{\ell\ell}$ [GeV]	$>$	12
$p_T^{\ell\ell}$ [GeV]	$>$	30
pf E_T^{miss} [GeV]	$>$	20
mpmet [GeV]	$>$	20
$m_T^{\ell\ell E_T^{miss}}$ [GeV]	$>$	40
$m_{\ell\ell}$ [GeV]	$<$	76
$\Delta R(\ell, \ell)$	$<$	2.5
Leading Jet cMVAv2	$>$	-0.5884

Table 8.8: Summary of the Top control region definition. The selection that changes with respect to the signal region is put in the bottom of the table.

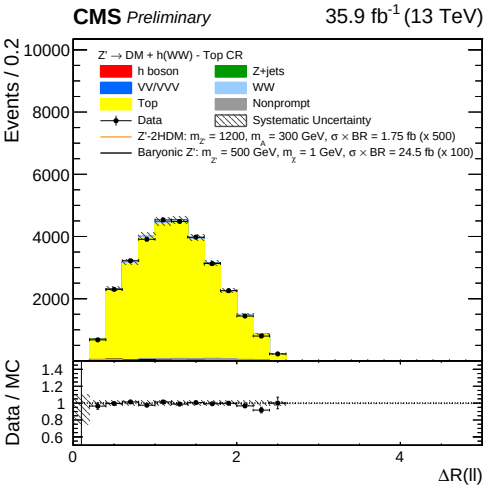
Figures 8.14 and 8.15 several kinematic distributions for the Top control region are shown, and a good agreement between data and simulation is observed.



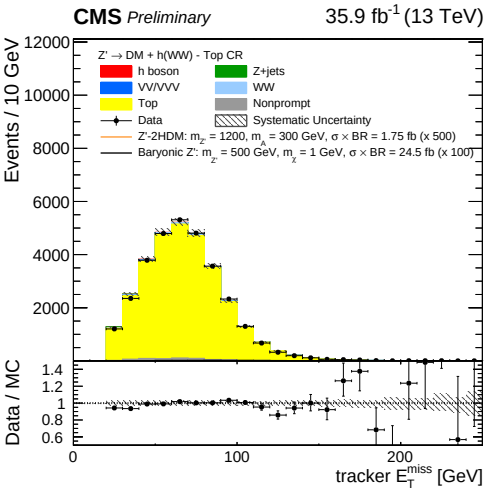
(a) $\Delta\phi(\ell,\ell)$.



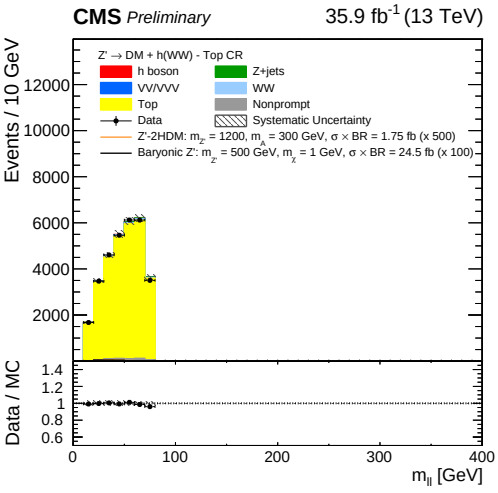
(b) $\mathrm{m}_{\mathrm{T}}^{\mathrm{W}_2}$.



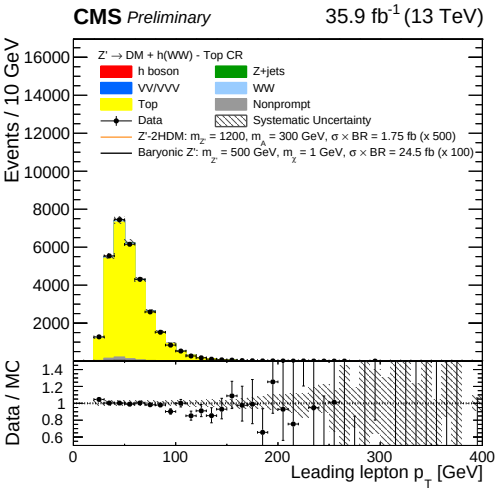
(c) $\Delta\mathrm{R}(\ell,\ell)$.



(d) track. $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$.

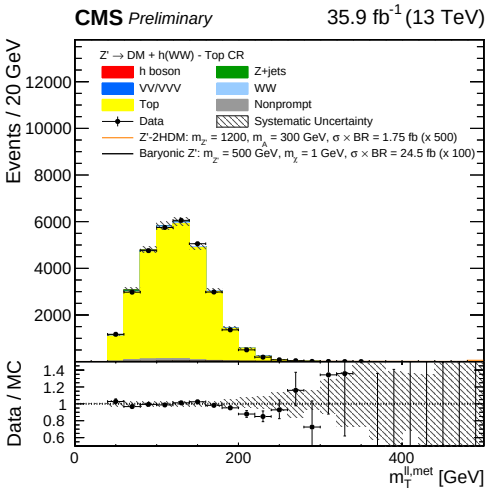


(e) $\mathrm{m}_{\ell\ell}$.

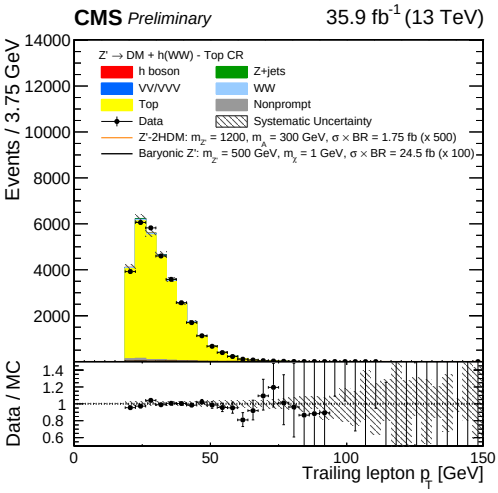


(f) leading lepton p_{T}.

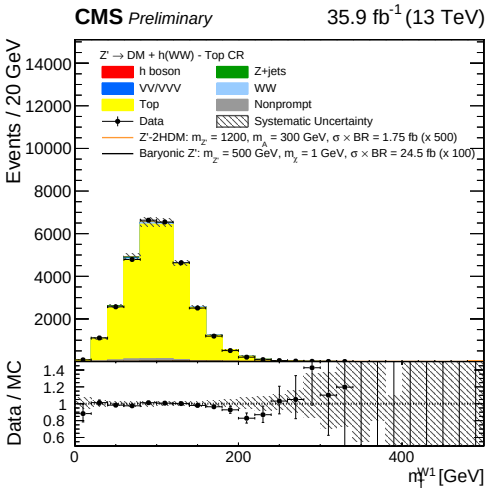
Figure 8.14: Control plots for several variables in the Top control region.



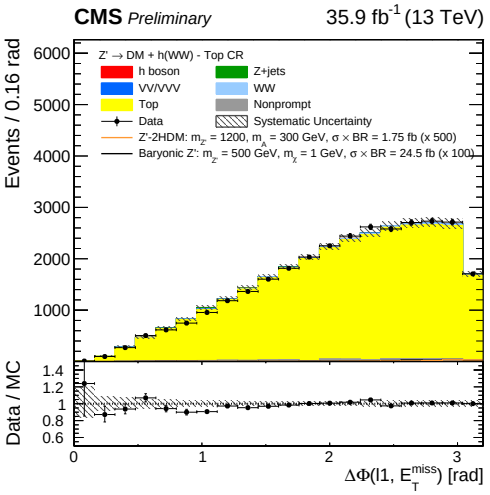
(a) $m_T^{\ell\ell E_T^{miss}}$.



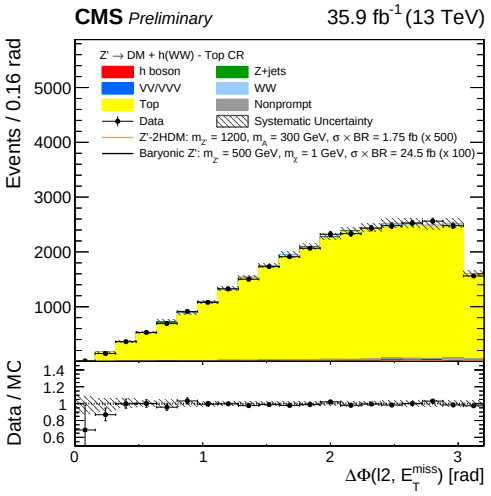
(b) trailing lepton p_T.



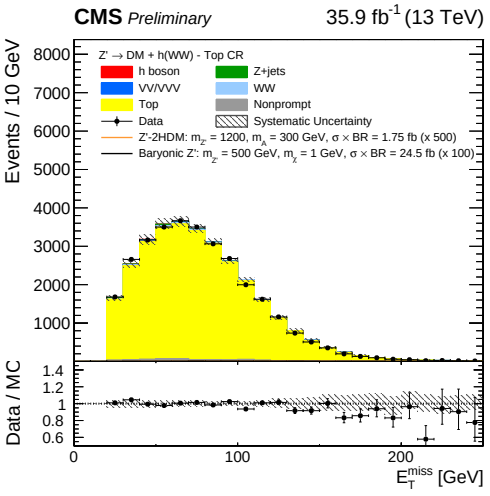
(c) $m_T^{W_1}$.



(d) $\Delta\phi(E_T^{miss}, \ell_1)$.



(e) $\Delta\phi(E_T^{miss}, \ell_2)$.



(f) E_T^{miss}.

Figure 8.15: Control plots for several variables in the Top control region.

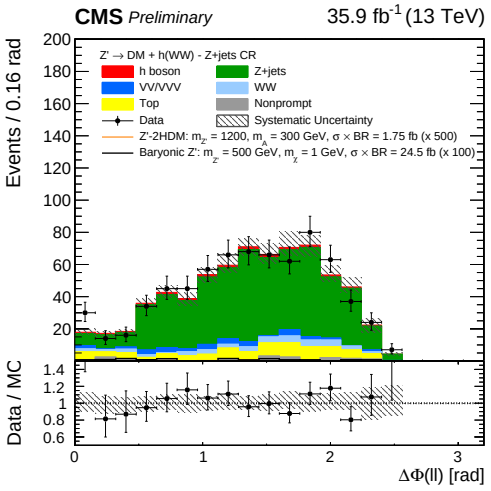
8.10 Drell-Yan Estimation

The $Z/\gamma^* \rightarrow \tau^+\tau^-$ process is a minor background for this search. It is simulated by using a dedicated MC sample, generated with aMC@NLO. A control region is defined to constrain the normalization of the process and to reduce its uncertainties. This control region is orthogonalized with respect to the signal region by requiring low $m_T^{\ell\ell E_T^{miss}}$, since the final-state leptons, coming from a lower mass resonance, which is not expected to be highly boosted, should provide a low transverse mass in the event. The definition of the $Z/\gamma^* \rightarrow \tau^+\tau^-$ control region can be found in Table 8.9. In Figures 8.16 and 8.17 several

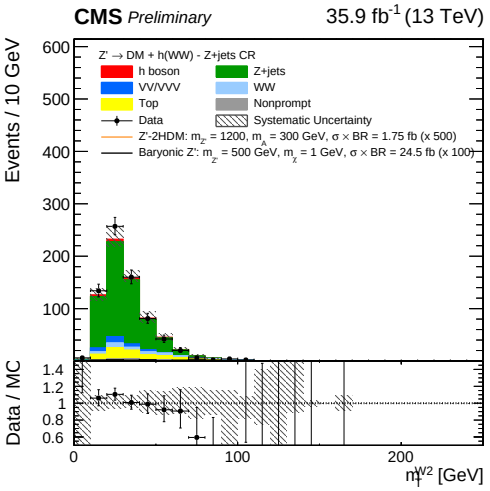
Variable	Selection	
$q_{\ell_1} \times q_{\ell_2}$	$<$	0
$p_T^{\ell_1}$ [GeV]	$>$	25
$p_T^{\ell_2}$ [GeV]	$>$	20
$p_T^{\ell_3}$ [GeV]	$<$	10
$m_{\ell\ell}$ [GeV]	$>$	12
$p_T^{\ell\ell}$ [GeV]	$>$	30
pf E_T^{miss} [GeV]	$>$	20
mpmet [GeV]	$>$	20
$m_{\ell\ell}$ [GeV]	$<$	76
$\Delta R(\ell, \ell)$	$<$	2.5
Jets cMVAv2	$<$	-0.5884
$m_T^{\ell\ell E_T^{miss}}$ [GeV]	$<$	40

Table 8.9: Summary of the Drell-Yan control region definition. The selection that changes with respect to the signal region is put in the bottom of the table.

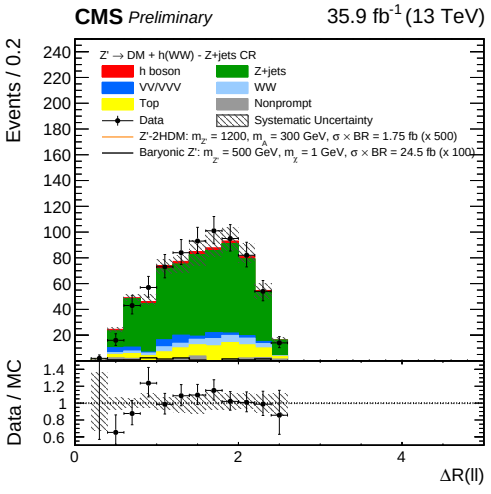
kinematic distributions for the Drell-Yan control region are shown, and a good agreement between data and simulation is observed.



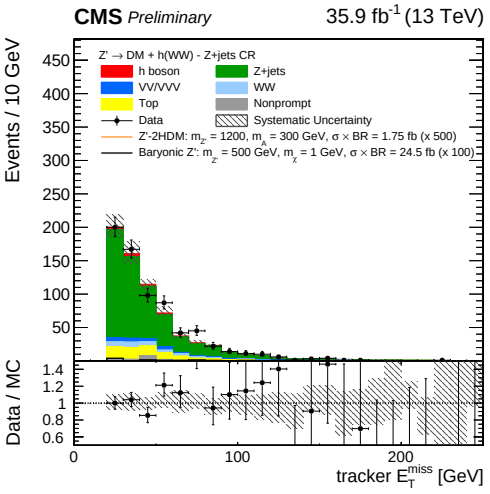
(a) $\Delta\phi(\ell,\ell)$.



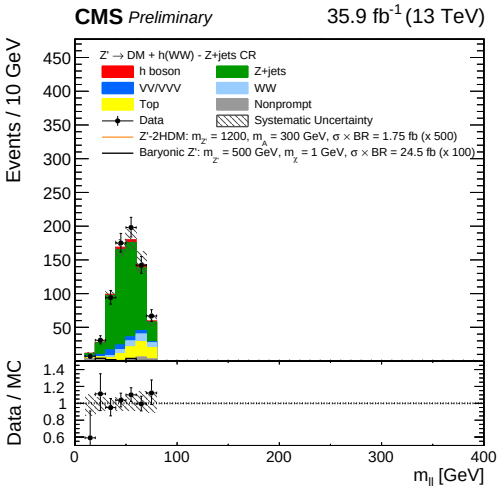
(b) $\mathrm{m_T^{W_2}}$.



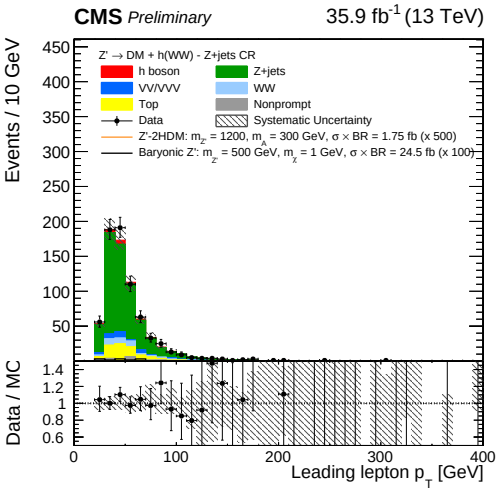
(c) $\Delta\mathrm{R}(\ell,\ell)$.



(d) track. $\mathrm{E_T^{miss}}$.

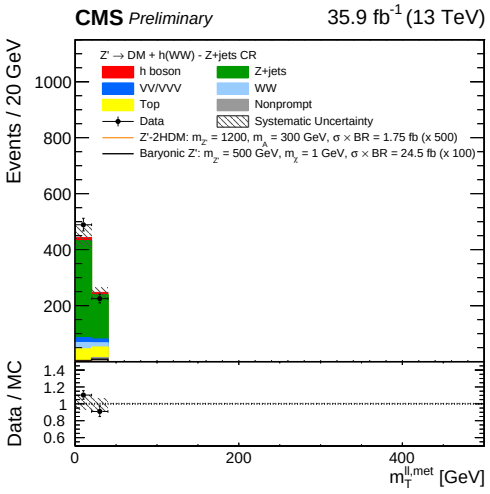


(e) $\mathrm{m}_{\ell\ell}$.

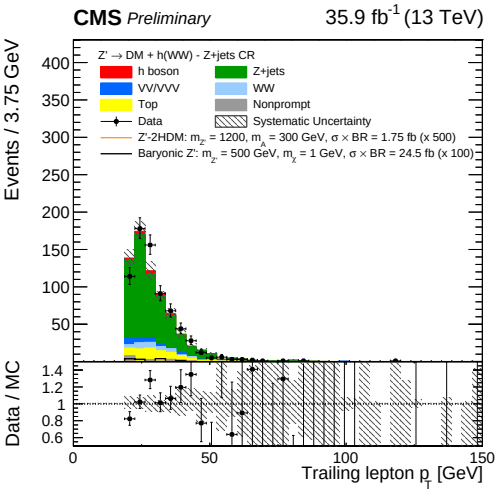


(f) leading lepton $\mathrm{p_T}$.

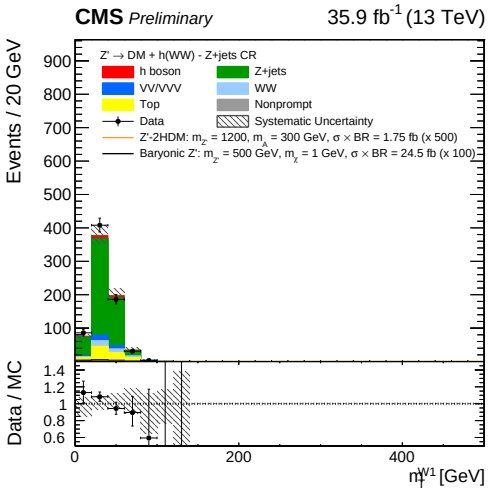
Figure 8.16: Control plots for several variables in the Drell-Yan control region.



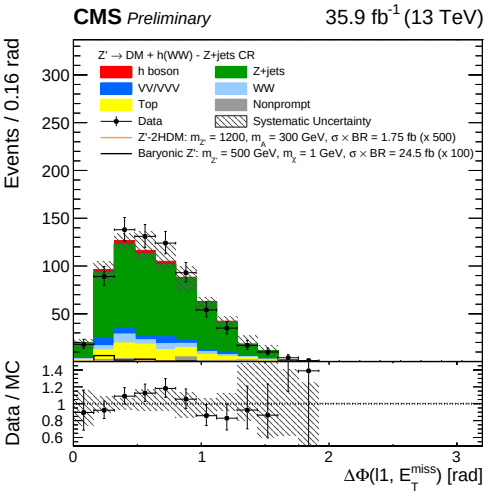
(a) $m_T^{\ell\ell E_T^{miss}}$.



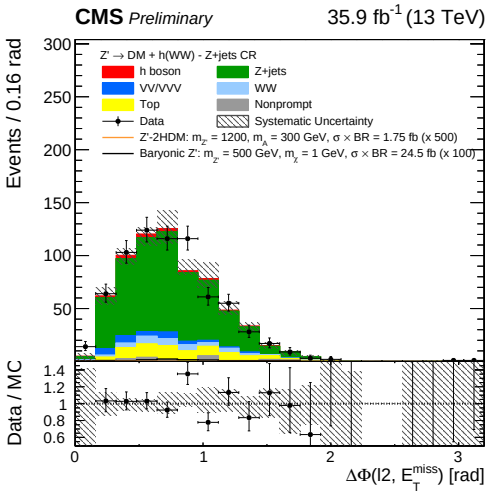
(b) trailing lepton p_T.



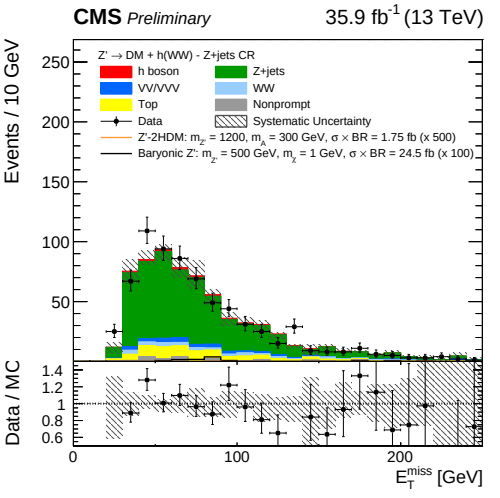
(c) $m_T^{W_1}$.



(d) $\Delta\phi(E_T^{miss}, \ell_1)$.



(e) $\Delta\phi(E_T^{miss}, \ell_2)$.



(f) E_T^{miss}.

Figure 8.17: Control plots for several variables in the Drell-Yan control region.

8.11 WZ and Wγ^* Estimation

The WZ and the Wγ^* processes involve the production of three leptons in the final state, so that they can contaminate the signal region if one of the three leptons is not detected or not reconstructed. They are both simulated by a single POWHEG MC sample, which includes the decay of a Z/γ^* with a lower di-lepton invariant mass threshold of 0.1 GeV. The normalization of the two processes is constrained using data in two different control regions. Since the large part of the γ^* events presents an invariant mass lower than 4 GeV, the data-driven normalization is performed by dividing the MC sample into two sub-samples, depending on the invariant mass of the two leptons coming from the decay of the Z/γ^*:

- WZ-like: $m_{\ell\ell} > 4$ GeV
- Wγ^*-like: 0.1 GeV $< m_{\ell\ell} < 4$ GeV

8.11.1 Normalization of WZ-like sub-sample

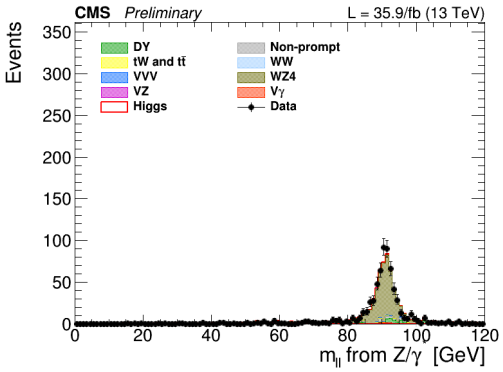
The normalization of the WZ-like sub-sample is measured using data by selecting three isolated, high p_T leptons. Two dedicated phase spaces have been defined, in order to select a region enriched in WZ events and to veto other processes: two muons and one electron final state, or two electrons and one muon final state. The same-flavour lepton pair is considered as coming from the decay of the Z boson, while the third lepton as coming from the W decay. To enhance the purity of the control region, some additional criteria are required:

- For the eeμ phase space:
  - two isolated Tight-ID electrons and one isolated Tight-ID muon;
  - two electrons with $p_T > 20$ GeV and $|\eta| < 2.1$, and a muon with $p_T > 25$ GeV and $|\eta| < 2.4$, to work on the plateau of the trigger efficiency for the ee path;
  - $E_T^{miss} > 30$ GeV to select events with a W boson;
- For the e$\mu\mu$ phase space:
  - two isolated Tight-ID muons and one isolated Tight-ID electron;
  - two muons with $p_T > 20$ GeV and $|\eta| < 2.4$ and an electron with $p_T > 20$ GeV and $|\eta| < 2.1$, to work on plateau of the trigger efficiency of double muon path;
  - $E_T^{miss} > 30$ GeV to select events with a W boson;

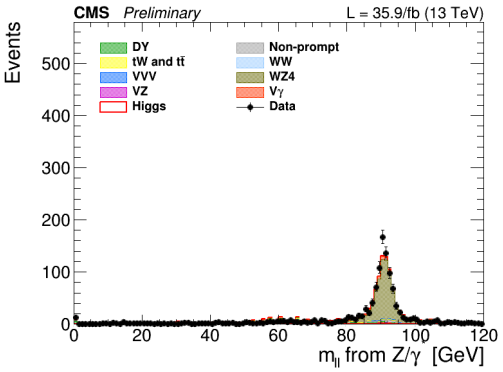
As shown in Figure 8.18, the normalization factor is estimated at the Z boson peak region ($80 < m_{\ell\ell} < 110$ GeV) on the same flavour, different charge lepton pair invariant mass spectrum. The scale factor extracted from this control region is 1.14 ± 0.18. Some kinematic distributions of the WZ-like phase space are shown in Figure 8.19, proving the good description of the process provided by the simulation.

8.11.2 Normalization of Wγ^*-like sub-sample

The normalization of the Wγ^*-like sub-sample is measured using data by selecting three tightly isolated leptons. In the calculation of the isolation of a lepton, the contribution of lepton tracks falling in its isolation cone is not taken into account (footprint removal). Two dedicated phase spaces have been used to measure the normalization factor: two muons, coming from the γ^* decay, and one high p_T muon or electron coming from the W decay. A Wγ^* enriched region is defined using the following selection criteria:

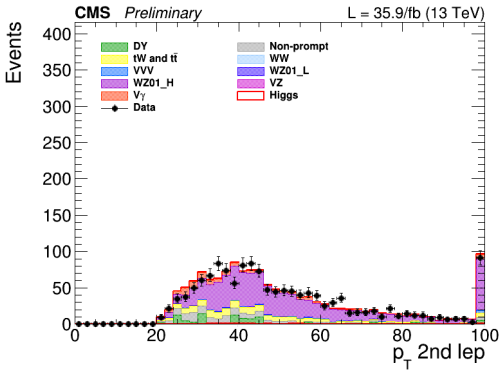


(a) di-electron mass

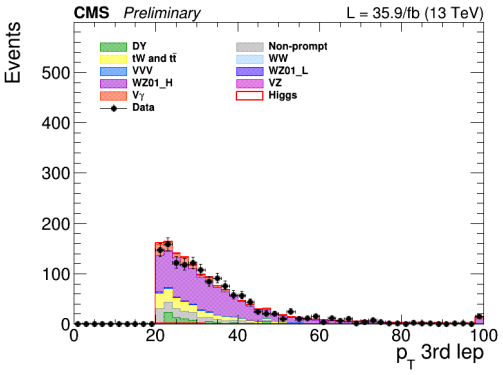


(b) di-muon mass

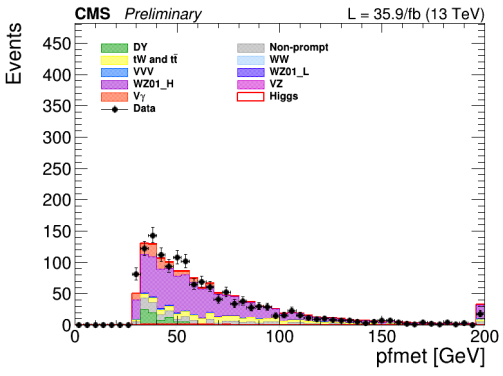
Figure 8.18: Same-flavour leptons invariant mass distributions in the WZ-like phase space. The region 80 GeV $< m_{\ell\ell} <$ 110 GeV is used to extract the the normalization scale factor.



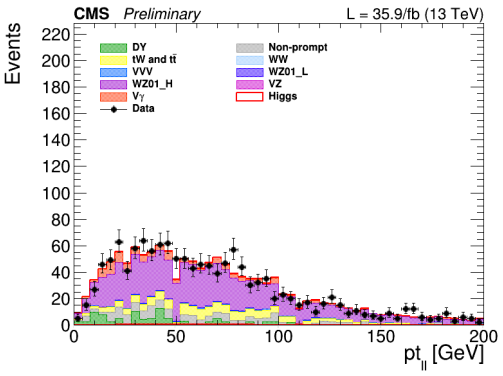
(a) Second lepton p_T.



(b) Third lepton p_T.



(c) E_T^{miss}.

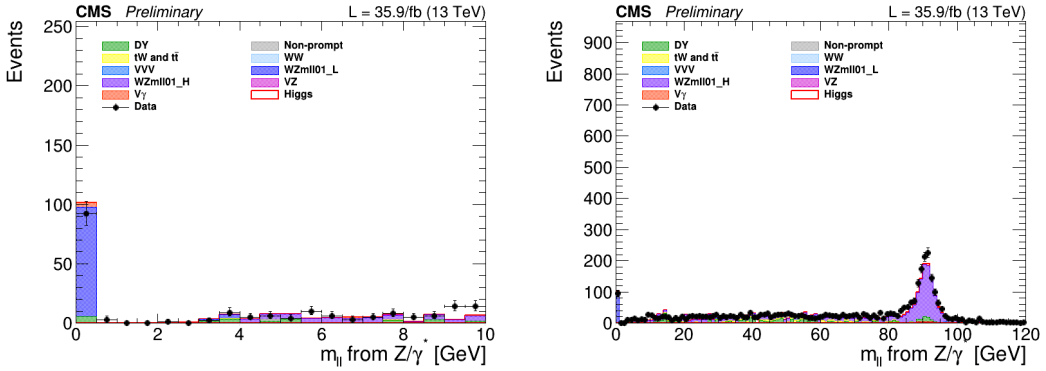


(d) $p_T^{\ell\ell}$.

Figure 8.19: Different kinematic variable distributions in the WZ enriched phase space.

- two isolated muons with $p_T > 8$ GeV, with footprint removal and opposite charge;
- invariant mass of the two muons incompatible with a J/ψ resonance (reject events with 2.2 GeV $< m_{\ell\ell} <$ 3.8 GeV);
- one Tight-ID electron with $p_T > 30$ GeV and $|\eta| < 2.1$, or one Tight-ID muon with $p_T > 30$ GeV and $|\eta| < 2.4$, to work on the plateau of the single lepton path trigger;
- mpmet > 25 GeV;

Since the muons used to define this phase space do not follow the object selection of the main analysis, the standard reconstruction and ID scale factors are not applied. To adjust the simulation to data, the Z peak is used instead: the MC prediction is adjusted to data by a normalization scale factor extracted in the 80 GeV $< m_{\ell\ell} <$110 GeV di-muon mass window, and then applied to the low-mass region. The scale factor is measured to be to be 1.24 $\pm$ 0.03. The normalization factor of the Wγ^*-like sub-sample is then measured in a low invariant mass region ($m_{\ell\ell} < 4$ GeV) to be 0.94 $\pm$ 0.24. Some kinematic distributions

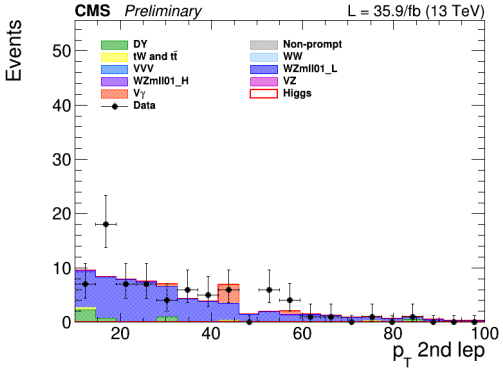


(a) di-muon mass (zoomed).

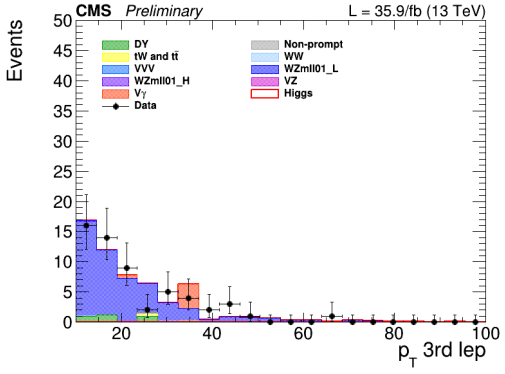
(b) di-muon mass.

Figure 8.20: Di-muon invariant mass distributions in the Wγ^* phase space. The very low mass region (on the left) is used to extract the normalization scale factor.

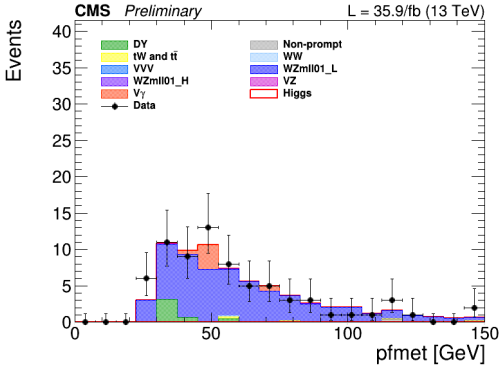
in the Wγ^* enriched phase space are plotted in Figure 8.21, showing a good description of the process by the simulated sample, considering the low statistics available.



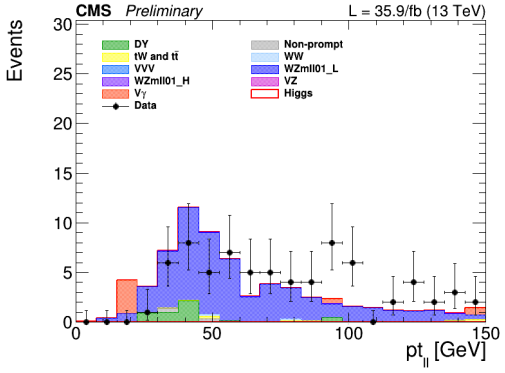
(a) Second lepton p_T.



(b) Third lepton p_T.



(c) E_T^{miss}.



(d) $p_T^{\ell\ell}$.

Figure 8.21: Different kinematic variable distributions in the Wγ^* enriched phase space.

8.12 Non-Prompt Leptons Estimation

The estimation of non-prompt leptons contamination in the signal region is fully data-driven and uses the so-called *fakeable object* method [139, 140]. The method consists in relating the properties of events with leptons reconstructed as Tight or Loose, to the properties of events with prompt or non-prompt leptons at generation level. In particular, through a correlation matrix, it is possible to estimate the probability for an event with one prompt lepton and one non-prompt lepton to be reconstructed as an event with two Tight leptons, and all the other possible combinations, as shown in Eq. 8.10.

$$\begin{bmatrix} \mathrm{TT} \\ \mathrm{TL} \\ \mathrm{LT} \\ \mathrm{LL} \end{bmatrix} = \begin{bmatrix} \mathrm{p}^2 & \mathrm{pf} & \mathrm{fp} & \mathrm{f}^2 \\ \mathrm{p}(1-\mathrm{p}) & \mathrm{p}(1-\mathrm{f}) & (1-\mathrm{f})\mathrm{p} & \mathrm{f}(1-\mathrm{f}) \\ (1-\mathrm{p})\mathrm{p} & (1-\mathrm{p})\mathrm{f} & \mathrm{f}(1-\mathrm{p}) & (1-\mathrm{f})\mathrm{f} \\ (1-\mathrm{p})^2 & (1-\mathrm{p})(1-\mathrm{f}) & (1-\mathrm{f})(1-\mathrm{p}) & (1-\mathrm{f})^2 \end{bmatrix} \begin{bmatrix} \mathrm{PP} \\ \mathrm{PF} \\ \mathrm{FP} \\ \mathrm{FF} \end{bmatrix} \tag{8.10}$$

On the left side of the equation the properties of the reconstructed event are listed: T stands for Tight lepton, L stands for Loose lepton, so that TT means an event with two Tight leptons, TL an event with the leading lepton passing Tight selection criteria and the trailing lepton passing Loose selection criteria, etc. The right side contains instead the properties of the *matrix-element* event: P stands for prompt lepton, F for *fake*, or non-prompt lepton, so that PP means an event with two prompt leptons, etc. The matrix connecting the reconstructed event to the generation-level event is built using the *prompt rate* (p) and the *fake rate* (f). The prompt rate is defined as the probability for a prompt lepton to be reconstructed as a Tight lepton, while the fake rate represents the probability for a non-prompt lepton to be reconstructed as a Tight lepton. The fake rate and the prompt rate are provided in bins of $\mathrm{p_T}$ and η of the leptons. They are then applied to a data sample selected with the analysis trigger paths, in which one lepton passes tight selection criteria and the other passes Loose selection criteria, but fail to pass Tight selections. This method thus provides both the kinematics and the normalization of the non-prompt lepton background. Since in data information at matrix-element level is not available, the prompt rate and the fake rate have to be defined through observable quantities, as explained in the next sections.

8.12.1 Fake Rate Estimation

In order to be estimated only through observable quantities, the fake rate is defined as the fraction of leptons that satisfy the Tight selection criteria, among those which satisfy the Loose selection criteria (*fakeable objects*):

$$\mathrm{f} = \frac{\text{number of Tight leptons}}{\text{number of Loose leptons}} \tag{8.11}$$

where Tight and Loose selections are described in Section 8.3 for muons and in Section 8.4 for electrons. The fake rate is measured in a QCD phase space, enriched in di-jet events, where a small presence of prompt leptons is expected. Such phase space is defined first of all by requiring events that have been accepted by the trigger paths listed in Table 8.2. The QCD enriched sample still may contain real leptons from W or Z leptonic decays that can bias the final lepton fake rate measurement. Events with muons from W decays are removed by requiring $\mathrm{E_T^{miss}} < 20$ GeV. The transverse mass built with the lepton and the $\mathrm{E_T^{miss}}$, corresponding to the W transverse mass, has to be lower than 20 GeV as well. Events with muons from Z decays are removed by requiring one and only one lepton.

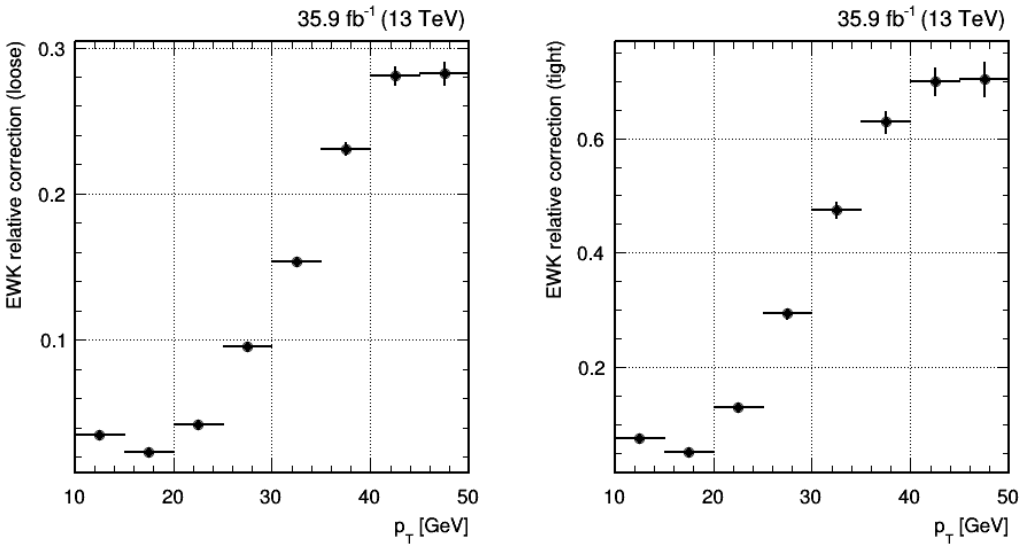
This procedure rejects most of the W and Z events while retaining almost all of the di-jet events in the QCD-enriched phase space. After applying this electroweak (EWK) veto, the remaining real lepton contamination from EWK events can still bias the fake rate, mainly at high p_T. To correct for this effect, the EWK contamination is estimated from two Monte Carlo samples, using the expected cross-section and the effective luminosity of the lepton triggers listed in Table 8.2, and subtracted from the data sample. In Figure 8.22 the relative EWK contaminations for events with one loose or one tight lepton are shown, separately for electrons and muons. The relative contamination increases with the lepton p_T and is larger when Tight leptons are selected. After the EWK subtraction, the fake rate can finally be computed. Both muon and electron candidates are required to be well separated from the leading jet of the event ($\Delta\phi(\ell, j) > 1$), such that they are almost back-to-back. For the computation of the electron fake rate, events in which the jet recoiling against the loosely identified electron has a E_T larger than 35 GeV are selected, while the muon fake rate considers events with the recoiling jet E_T larger than 25 GeV. The results are shown in Figure 8.23, separately for electrons and muons.

8.12.2 Prompt Rate Estimation

The estimation of the prompt rate is obtained with the Tag and Probe method, in a region enriched in prompt leptons, under the peak of the $Z \to \ell^+\ell^-$ resonance. The prompt rates for electrons and muons are shown in Figure 8.24.

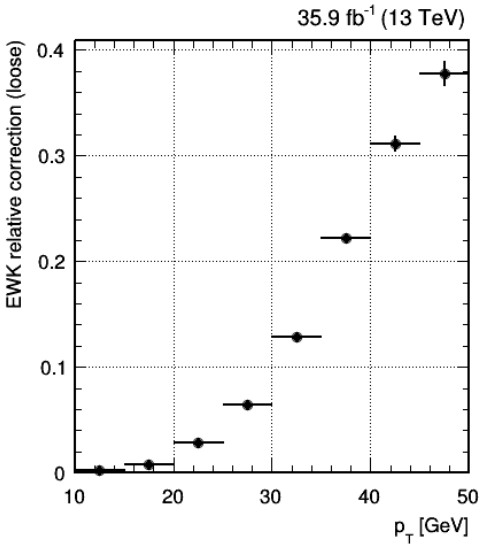
8.12.3 Validation of the Method

The fundamental assumption of the fakeable object method is that, once the lepton kinematics is taken into account, the same universal fake rate measured in a specific phase space can be applied to a different phase space to evaluate the non-prompt contamination there. To test this hypothesis and verify that the method is able to reproduce the kinematic of processes that produce non-prompt leptons in the signal region, two validation regions have been defined. Both of them require two leptons of the same charge, since the non-prompt lepton charge is not expected to depend on the charge of the prompt lepton in the event, and much fewer SM processes produce same-sign lepton pairs than opposite-sign lepton pairs so that the contribution of non-prompt leptons can be more easily seen. The first validation region is defined as the signal region, with just the inversion of the lepton charge selection. The second one is instead dedicated to the validation of the method in a phase space enriched in b-quarks, and mimics the selections of the top control region, but with two same-sign leptons. The full selections defining the two validation regions are listed in Table 8.10. Kinematic distributions of the validation regions are shown in Figures 8.25 and 8.26 for the b-veto region, and in Figures 8.27 and 8.28 for the b-tag region. A general good agreement with data is found, ensuring the validity of the method, and the maximum discrepancy between data and prediction ($\sim 30\%$) is taken as its systematic uncertainty.

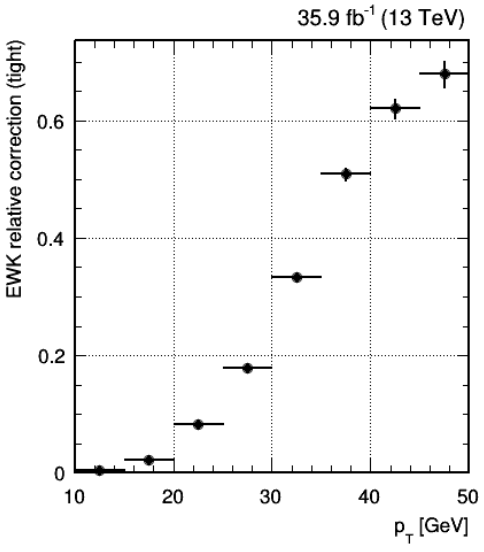


(a) Loose electrons.

(b) Tight electrons.

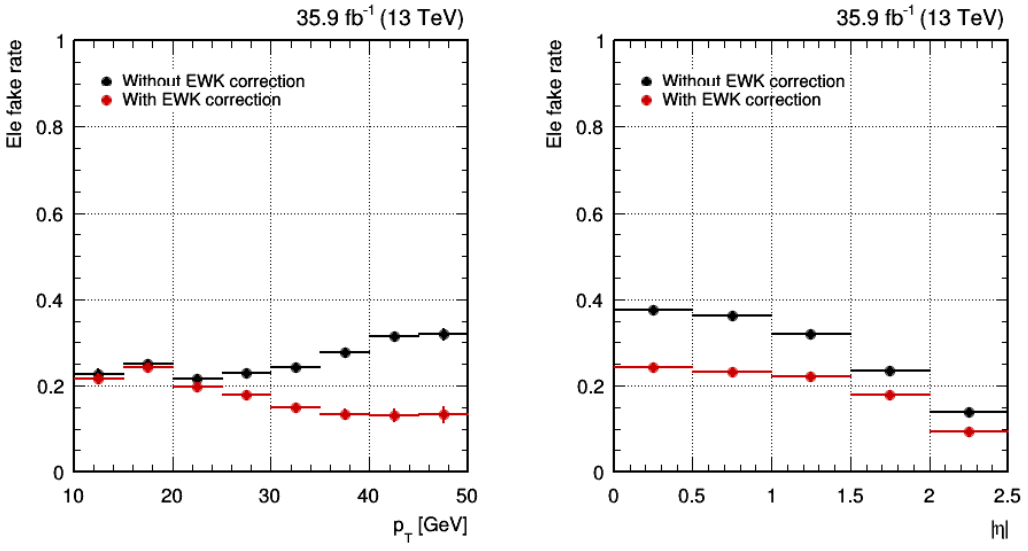


(c) Loose muons.



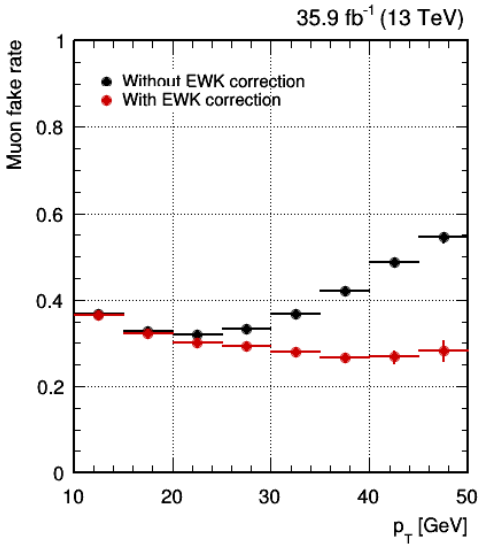
(d) Tight muons.

Figure 8.22: Relative EWK contamination for the QCD-enriched phase space when events with one loose lepton (left) or one tight electron (right) are selected, separately for electrons (top) and muons (bottom). The contamination increases with the lepton p_T and is significantly higher when Tight leptons are selected.

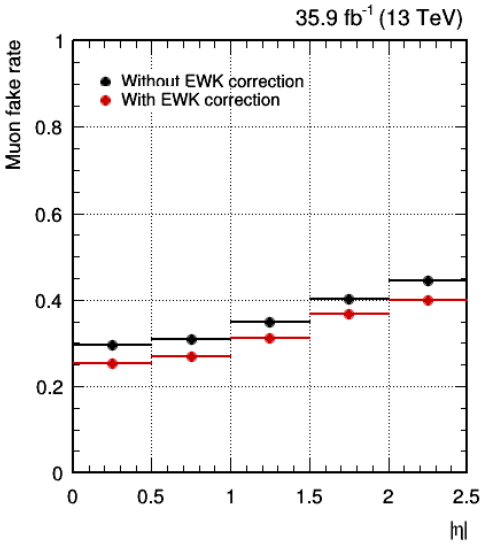


(a) Electron p_T.

(b) Electron η.

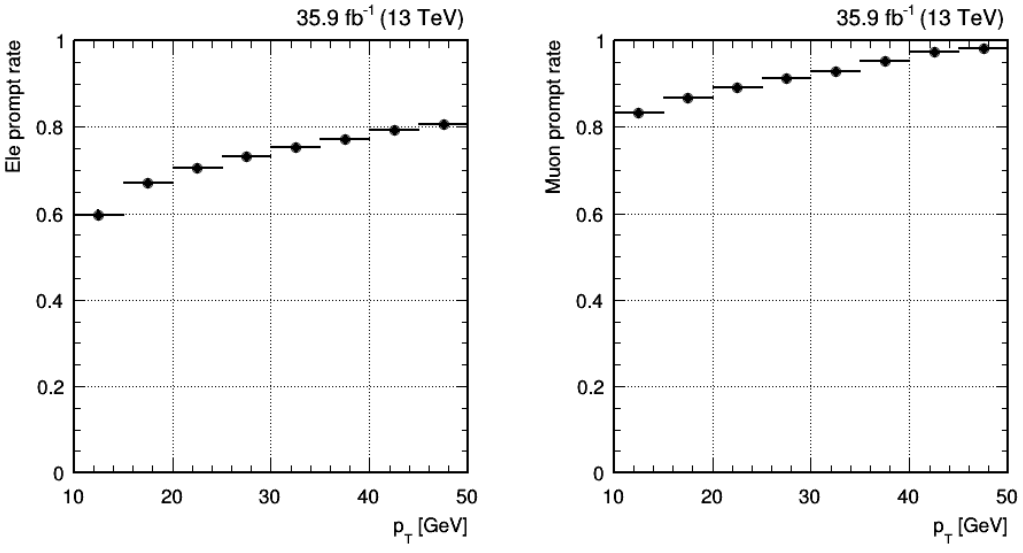


(c) Muon p_T.



(d) Muon η.

Figure 8.23: Fake rate as a function of the loose lepton p_T (left) and η (right), before and after EWK correction (in black and red, respectively), separately for electrons (top) and muons (bottom).



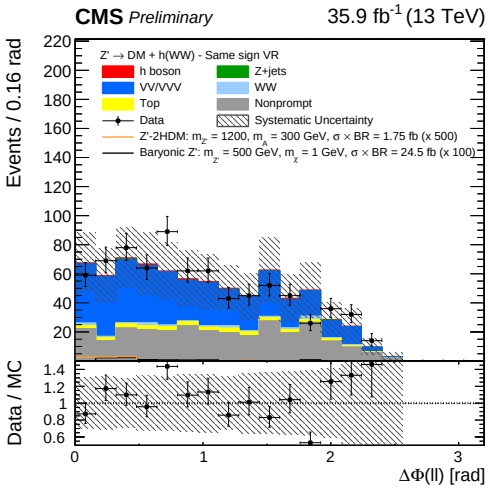
(a) Electrons.

(b) Muons.

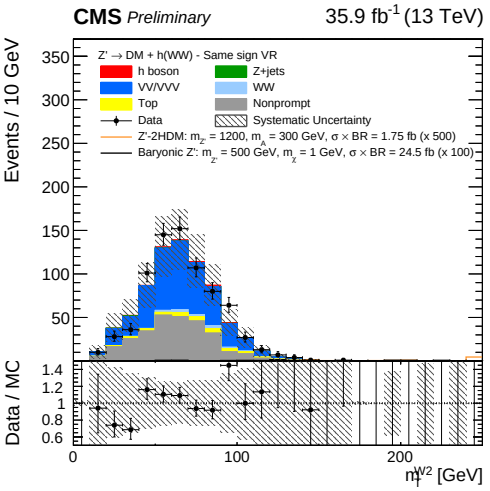
Figure 8.24: Prompt rate as a function of the loose lepton p_T for electrons (left) and muons (right).

Variable	B-Veto		B-tag	
$p_T^{\ell_1}$ [GeV]	$>$	25	$>$	25
$p_T^{\ell_2}$ [GeV]	$>$	20	$>$	20
$p_T^{\ell_3}$ [GeV]	$<$	10	$<$	10
$m_{\ell\ell}$ [GeV]	$>$	12	$>$	12
$p_T^{\ell\ell}$ [GeV]	$>$	30	$>$	30
pf E_T^{miss} [GeV]	$>$	20	$>$	20
mpmet [GeV]	$>$	20	$>$	20
$m_T^{\ell\ell E_T^{miss}}$ [GeV]	$>$	40	$>$	40
Jets cMVAv2	$<$	-0.5884	$<$	-0.5884
$m_{\ell\ell}$ [GeV]	$>$	76	$>$	76
$\Delta R(\ell, \ell)$		-		-
$q_{\ell_1} \times q_{\ell_2}$	$>$	0	$>$	0
Jets cMVAv2	$<$	-0.5884		-
Leading Jet cMVAv2		-	$>$	-0.5884

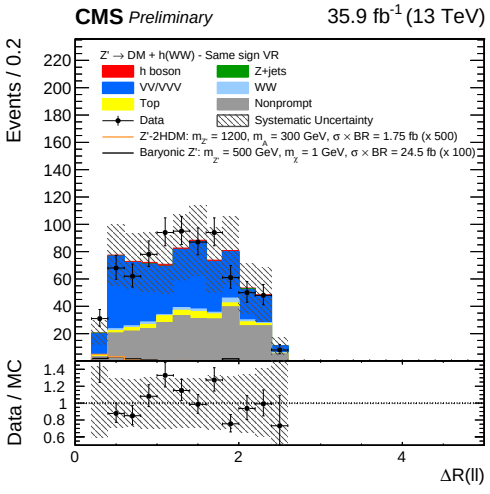
Table 8.10: Summary of the two same-sign validation regions definitions. The selections that change with respect to the signal region are put at the bottom of the table.



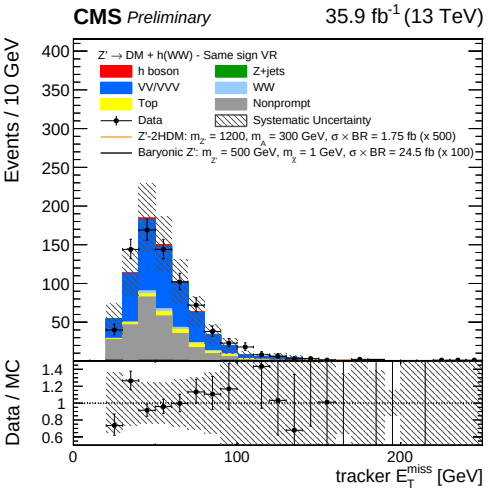
(a) $\Delta\phi(\ell,\ell)$.



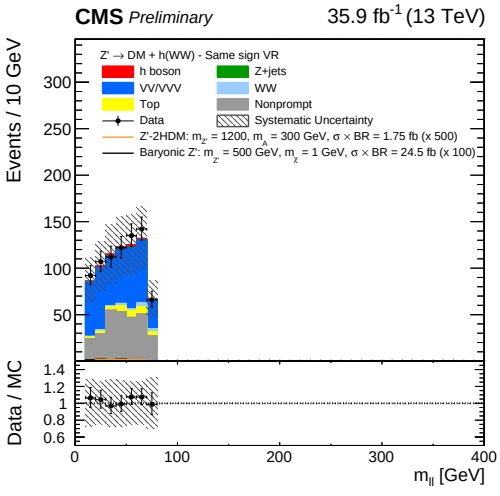
(b) $\mathrm{m}_{\mathrm{T}}^{\mathrm{W}_2}$.



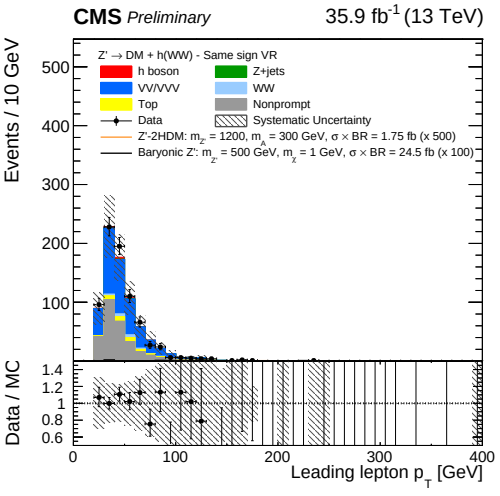
(c) $\Delta\mathrm{R}(\ell,\ell)$.



(d) track. $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$.

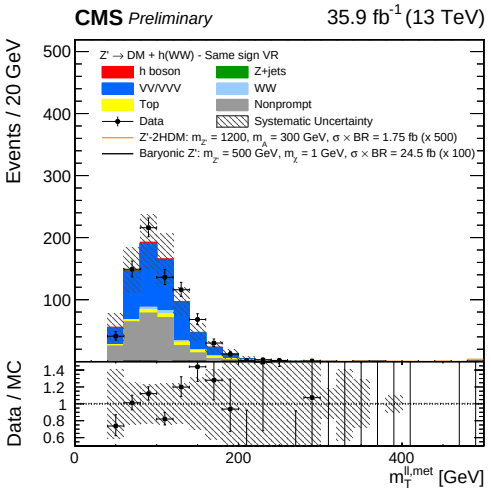


(e) $\mathrm{m}_{\ell\ell}$.

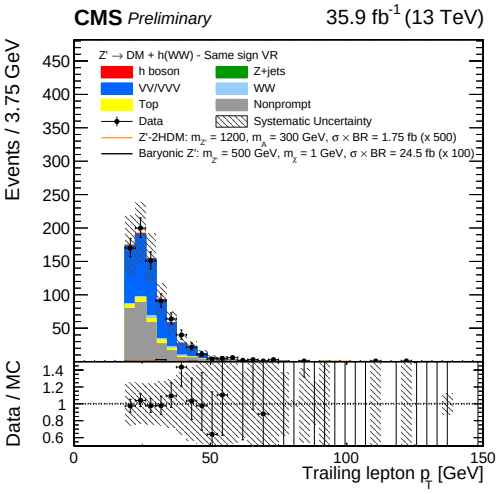


(f) leading lepton p_{T}.

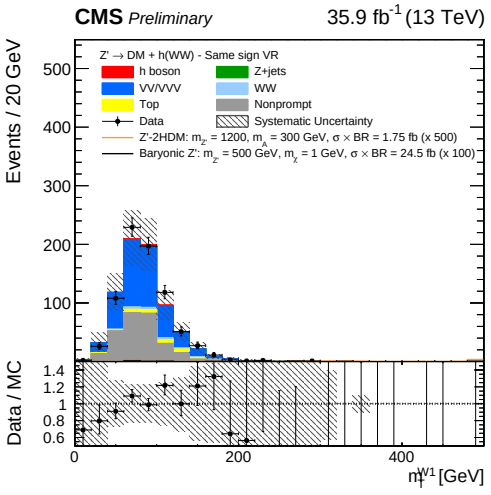
Figure 8.25: Control plots for several variables in the b-veto same-sign validation region.



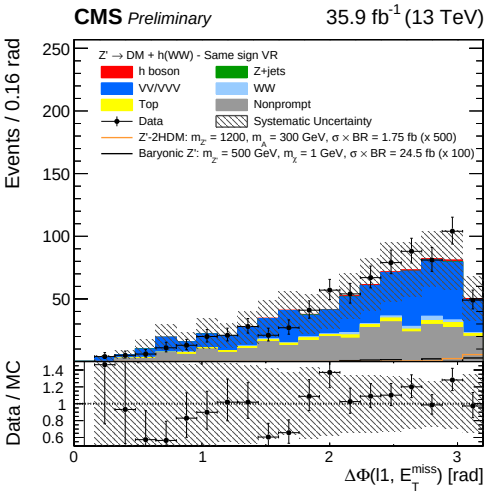
(a) $m_T^{\ell\ell E_T^{miss}}$.



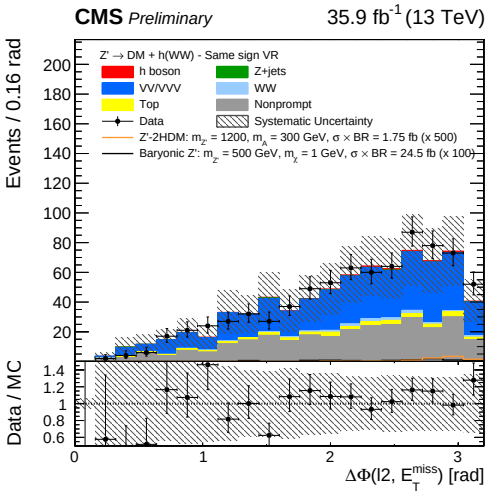
(b) trailing lepton p_T.



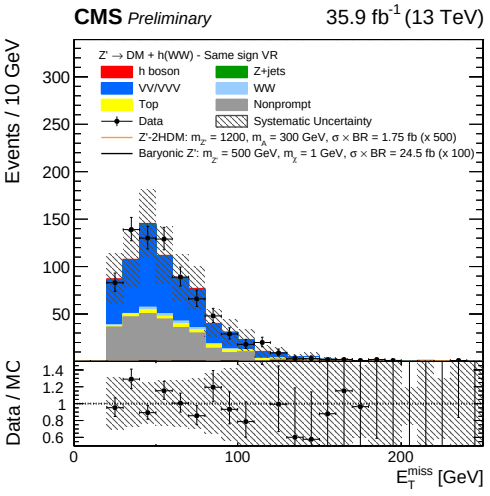
(c) $m_T^{W_1}$.



(d) $\Delta\phi(E_T^{miss}, \ell_1)$.

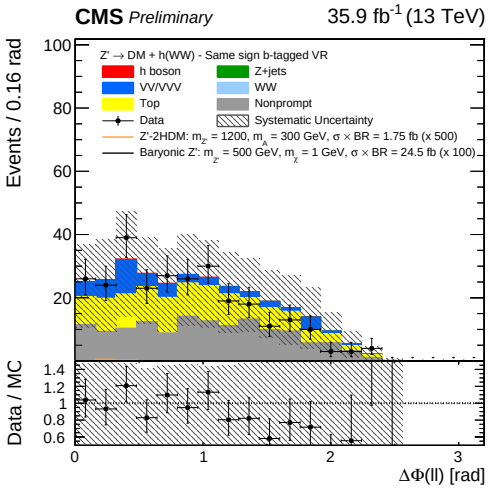


(e) $\Delta\phi(E_T^{miss}, \ell_2)$.

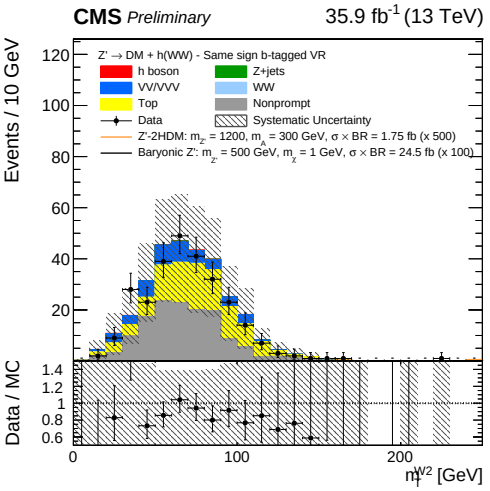


(f) E_T^{miss}.

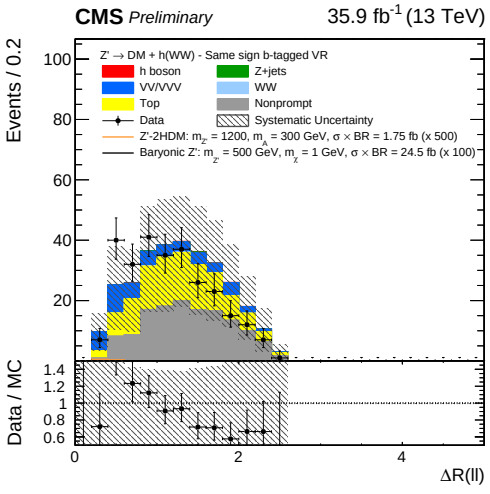
Figure 8.26: Control plots for several variables in the b-veto same-sign validation region.



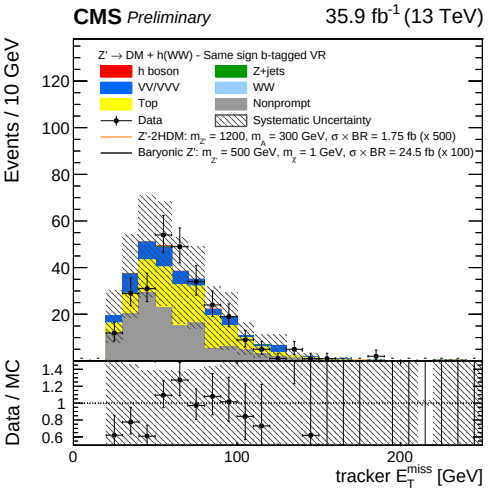
(a) $\Delta\phi(\ell,\ell)$.



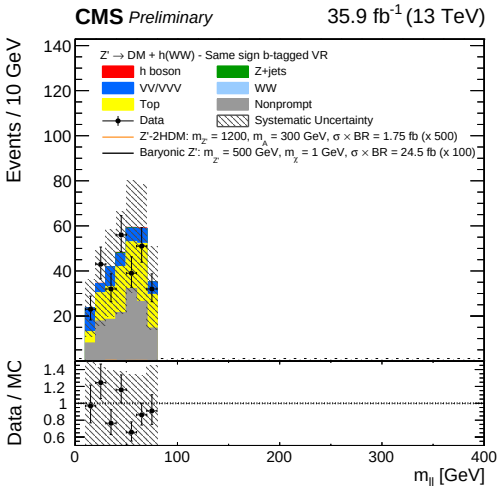
(b) $\mathrm{m}_{\mathrm{T}}^{\mathrm{W}_2}$.



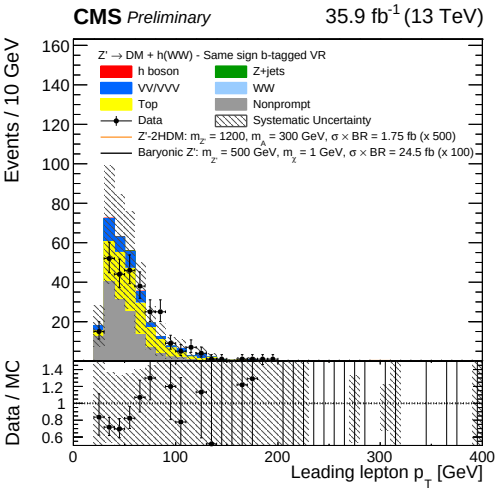
(c) $\Delta\mathrm{R}(\ell,\ell)$.



(d) track. $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$.

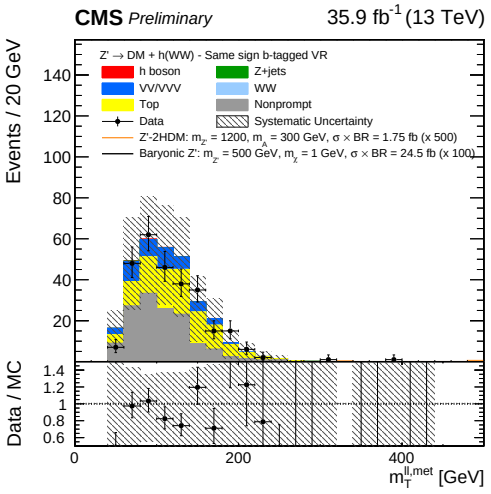


(e) $\mathrm{m}_{\ell\ell}$.

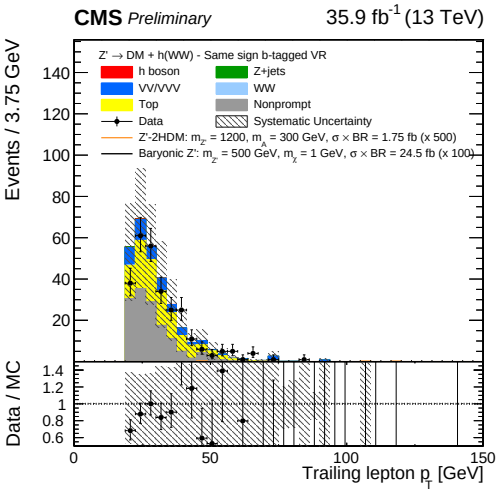


(f) leading lepton p_{T}.

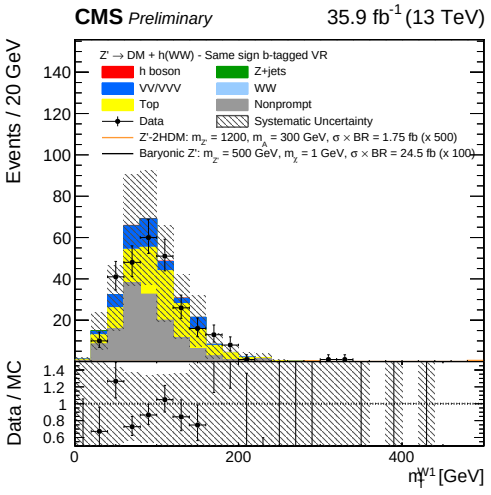
Figure 8.27: Control plots for several variables in the b-tag same-sign validation region.



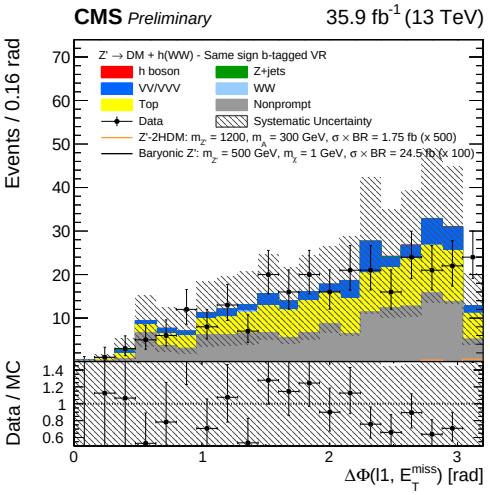
(a) $m_T^{\ell\ell E_T^{miss}}$.



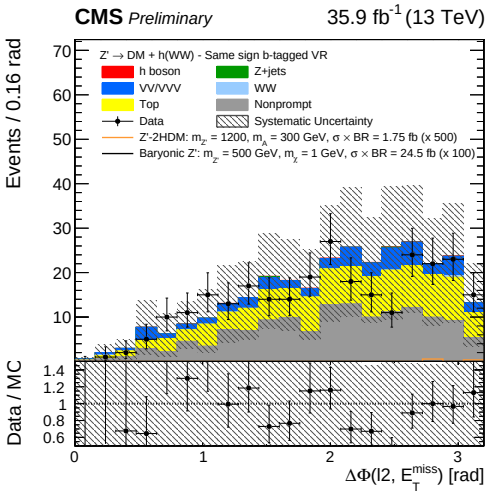
(b) trailing lepton p_T.



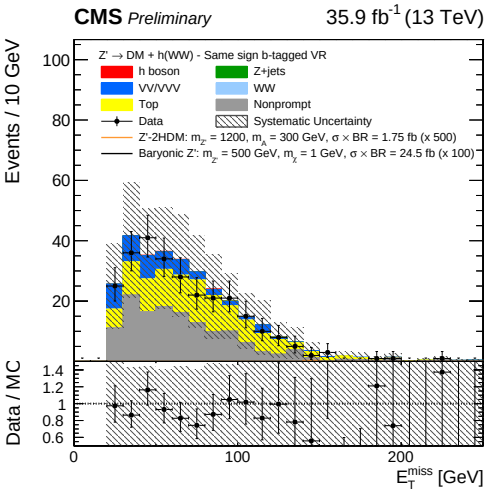
(c) $m_T^{W_1}$.



(d) $\Delta\phi(E_T^{miss}, \ell_1)$.



(e) $\Delta\phi(E_T^{miss}, \ell_2)$.



(f) E_T^{miss}.

Figure 8.28: Control plots for several variables in the b-tag same-sign validation region.

8.13 Other Backgrounds Estimation

The estimation of the remaining backgrounds, for which it is not possible to define a control region, or represent a small source of contamination for the signal region, relies on the MC simulation. In particular, the kinematics of all the Higgs boson production modes is simulated at NLO with POWHEG, with the exception of the bbh production, which is simulated with aMC@NLO. All the cross-sections have been taken from the most recent NNLO calculations [89]. Also the di-boson and tri-boson processes, excluding the already discussed WW, WZ, and Wγ^*, are estimated directly using MC prediction, normalizing the cross-sections to the theoretical calculations. A summary of all the MC samples used in this document for the backgrounds estimation is presented in Table 8.11 for Higgs processes, in Table 8.12 for WW and $t\bar{t}$, and in Table 8.13 for the remaining backgrounds.

process	dataset name	events
ggh	GluGluHToWWTo2L2Nu_M125_13TeV_powheg_JHUgen_pythia8	500K
VBF	VBFHToWWTo2L2Nu_M125_13TeV_powheg_JHUgen_pythia8	500K
W^+h	HWplusJ_HToWW_M125_13TeV_powheg_pythia8	300K
W^-h	HWminusJ_HToWW_M125_13TeV_powheg_pythia8	300K
Zh	HZJ_HToWWTo2L2Nu_M125_13TeV_powheg_pythia8	3M
ggZh	GluGluZH_HToWWTo2L2Nu_M125_13TeV_powheg_pythia8	500K
bbh	bbHToWWTo2L2Nu_M-125_4FS_yb2_13TeV_amcatnlo	750K
	bbHToWWTo2L2Nu_M-125_4FS_ybyt_13TeV_amcatnlo	750K
h$\to\tau\tau$	GluGluHToTauTau_M125_13TeV_powheg_pythia8	1.5M
	VBFHToTauTau_M125_13TeV_powheg_pythia8	1.5M
	ZHToTauTau_M125_13TeV_powheg_pythia8	800K
	WplusHToTauTau_M125_13TeV_powheg_pythia8	500K
	WminusHToTauTau_M125_13TeV_powheg_pythia8	500K

Table 8.11: Reference Higgs background samples used in the analysis.

process	dataset name	events
$t\bar{t} \to WWb\bar{b} \to 2\ell 2\nu b\bar{b}$	TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	80M
$q\bar{q} \to WW \to 2\ell 2\nu$	WWTo2L2Nu_13TeV-powheg	2000K
$gg \to WW \to 2\ell 2\nu$	GluGluWWTo2L2Nu_MCFM_13TeV	500K

Table 8.12: Simulated samples for $t\bar{t}$ and WW production.

process	dataset name
single top	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1
	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1
	ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1
Drell-Yan	DYJetsToTauTau_ForcedMuEleDecay_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
multibosons	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
	WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8
	ZZTo2L2Nu_13TeV_powheg_pythia8
	WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8
	WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
	WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
	ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8
$W\gamma$	WGToLNuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
$W\gamma^* \to \ell\nu\mu\mu$	WGstarToLNuMuMu_012Jets_13TeV-madgraph
$W\gamma^* \to \ell\nu ee$	WGstarToLNuEE_012Jets_13TeV-madgraph

Table 8.13: Simulated samples for other backgrounds used in the analysis.

Chapter 9

Signal Extraction

Condotti da fragili desideri
Tra puro movimento immoto

CCCP, *Trafitto*

As already seen in Section 8.7, the signal region suffers background contamination, and on the other hand, several kinematic variables indicate that it is possible to enhance the discrimination between signal and background, even if none of them can do it alone. A multivariate analysis (MVA) technique has thus been exploited, in particular through *boosted decision trees* (BDTs), to perform the signal extraction. Two different sets of BDT have been trained, one for each signal model. No selections have been put on the MVA output, instead, a binned likelihood fit to its shape has been performed, in order to avoid losing signal events, and to further constrain the backgrounds normalization in the signal region, using the bins of the distributions where no significant signal contribution is expected. In this chapter, the details of the BDT training and of the shape analysis, including the uncertainties considered, will be described.

9.1 Multivariate Analysis

To exploit all the several kinematic variables able to discriminate between signal and main backgrounds and their correlations, two different BDTs have been trained for the two signal models considered. The BDT output discriminant returns values between -1 and +1 for each event, where a score close to +1 indicates that the event is signal-like, and a score close to -1 means that the event is background-like. For both models, the same variables have been used:

- Transverse masses: $m_T^{\ell\ell E_T^{miss}}$, $m_T^{W_1}$, $m_T^{W_2}$.
- Leptons momenta: $p_T^{\ell\ell}$, $p_T^{\ell_1}$, $p_T^{\ell_2}$.
- Missing energies: PF E_T^{miss}, tracker E_T^{miss}, mpmet.
- Angular variables: $\Delta R(\ell,\ell)$, $\Delta\phi(\ell,\ell)$, $\Delta\phi(E_T^{miss},\ell_1)$, $\Delta\phi(E_T^{miss},\ell_2)$
- Dilepton invariant mass: $m_{\ell\ell}$

The training is performed after applying the signal region selection (explained in Section 8.7), since here very few signal events are rejected, and there is sufficient statistics to produce a sensible training. Different training algorithms and settings have been tested in order to get the best signal versus background separation and a low level of overtraining. All the studies performed are listed in Appendix B. According to the results of such studies, the training parameters chosen are the same for both models and are listed in Table 9.1. Since several signal mass points are inspected for each model, the BDTs have

Adaptive Boost
NTrees = 500
MinNodeSize = 0.5%
BoostType = AdaBoost
AdaBoostBeta = 0.1
nCuts = 500
MaxDepth = 2

Table 9.1: Summary of BDT training options used in the analysis. The parameters chosen are the same for the two models inspected.

been optimized in order to better discriminate those points for which some sensitivity is expected to be accessible with this analysis, by using only those points for the training, depending on the model. For the masses selected, the BDTs are trained in order to give a response as independent as possible of the specific sample. This is obtained by weighting the samples according to the specific cross-section of each mass point, and then correcting the weights as follows:

- The weight is divided by the **cross-section** of the mass point. This makes the BDT insensitive to the cross section of the samples. If such weight is not applied, mass points with larger cross sections would be seen by the BDT as more interesting, introducing a bias.
- The weight is divided by the **number of MC events** passing the signal region selections, in order to reduce the dependence of the BDT on the kinematics of the specific mass point. A mass point with a heavier mediator would, in fact, produce harder leptons and larger E_T^{miss}, so that more MC events would pass the signal region selections and would receive more importance from the BDT.

The main backgrounds producing two prompt leptons and E_T^{miss} in the final state, namely SM Higgs, WW, and $t\bar{t}$ are considered in the training, weighted according to their cross sections. The details of the training for the two models are given in the following of this section.

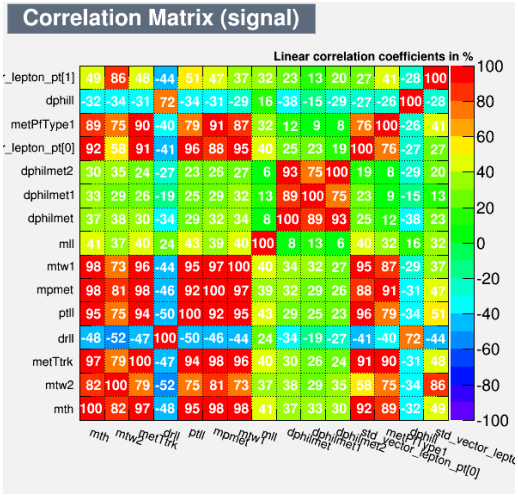
9.1.1 Training for Z'-2HDM Model

All the signal mass points with m_A = 300 GeV are considered for the training since for heavier masses of the pseudoscalar A the sensitivity of the analysis rapidly drops. The linear correlations among the variables used for the BDT training are shown in the top part of Figure 9.1. Here it can be seen that even if some variables are strongly correlated for the signal, the correlation is lower for the background, so they can add some discriminating power. In the bottom part of Figure 9.1, the Receiver Operating Characteristic

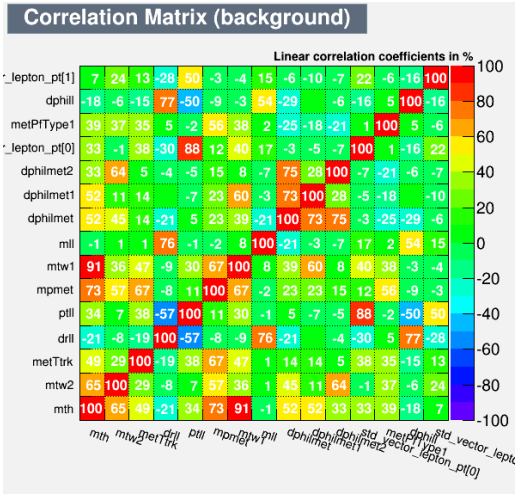
(ROC) curve, namely the plot showing the background rejection obtainable, given a selected signal efficiency and the overtraining and performance plots are shown. In particular, the last plot shows the good discrimination reached between the signal (in blue) and the background (in red), and the agreement between the training and test samples can be appreciated, as expected in case of low level of overtraining. The variables used for the BDT training are shown in Figure 9.2 and their ranking, namely the relative contribution to the discrimination power, are presented in Table 9.2. Finally, in Figure 9.3, the BDT output is plotted for data and MC, in the signal region and in the WW, Top, and Drell-Yan control regions, and a good agreement between data and simulation is observed.

Rank	Variable	Separation
1	$m_T^{\ell\ell E_T^{miss}}$	7.067×10^{-1}
3	mpmet	6.615×10^{-1}
2	$m_T^{W_1}$	6.598×10^{-1}
4	$m_T^{W_2}$	6.479×10^{-1}
5	track. E_T^{miss}	6.372×10^{-1}
6	$p_T^{\ell\ell}$	5.543×10^{-1}
7	$\Delta R(\ell, \ell)$	5.448×10^{-1}
8	PF E_T^{miss}	5.301×10^{-1}
9	$\min[\Delta\phi(E_T^{miss}, \ell_1), \Delta\phi(E_T^{miss}, \ell_2)\ ]$	4.769×10^{-1}
10	$p_T^{\ell_1}$	4.761×10^{-1}
11	$\Delta\phi(\ell, \ell)$	3.515×10^{-1}
12	$\Delta\phi(E_T^{miss}, \ell_1)$	3.210×10^{-1}
13	$\Delta\phi(E_T^{miss}, \ell_2)$	2.820×10^{-1}
14	$p_T^{\ell_2}$	2.643×10^{-1}
15	$m_{\ell\ell}$	7.003×10^{-2}

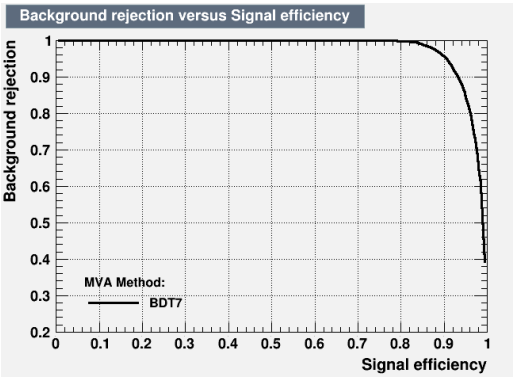
Table 9.2: BDT input variables ranking for Z'-2HDM model.



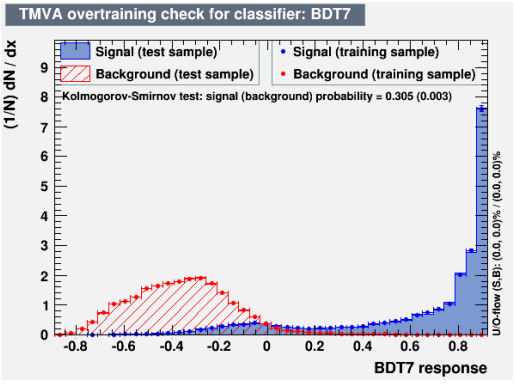
(a) Signal correlation matrix.



(b) Background correlation matrix.

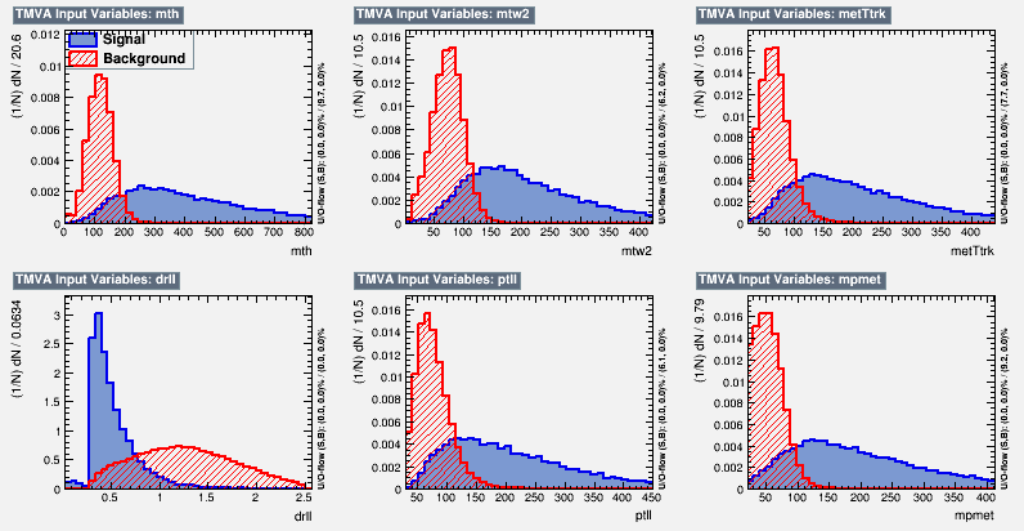


(c) ROC curve.

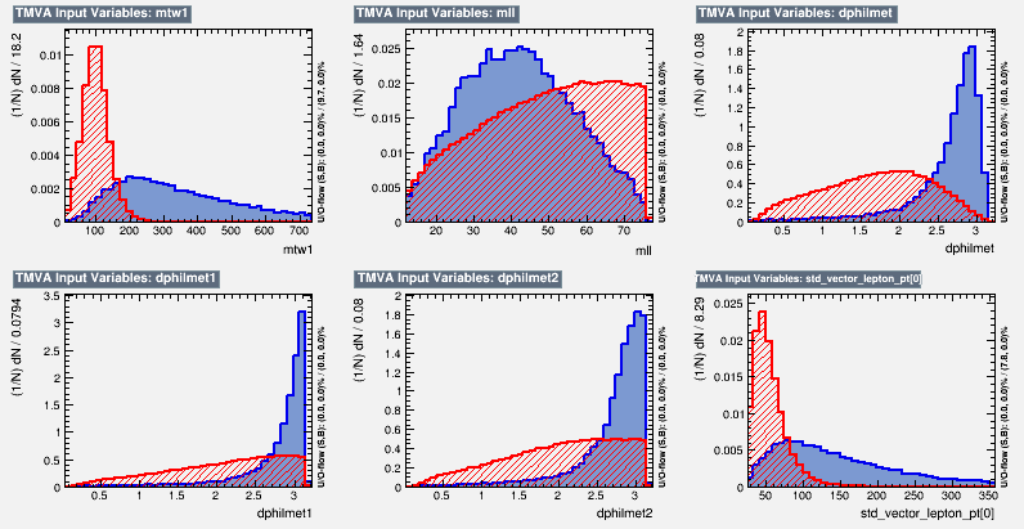


(d) Overtraining plot.

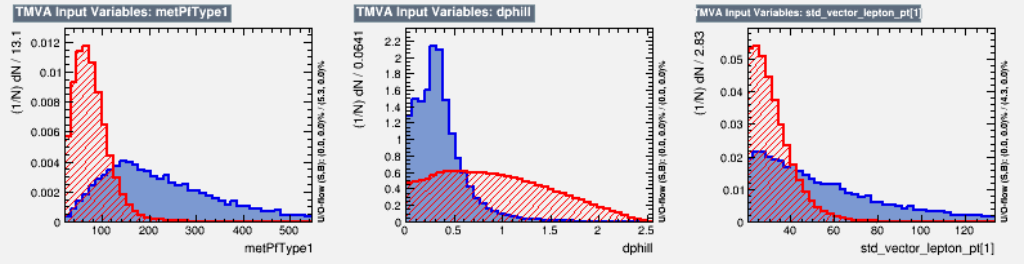
Figure 9.1: Correlation matrix for MVA used in the analysis (top), shown separately for the Z'-2HDM signal (left) and the background (right). The ROC curve (bottom left) and the overtraining and performance plots (bottom right) are also shown.



(a)

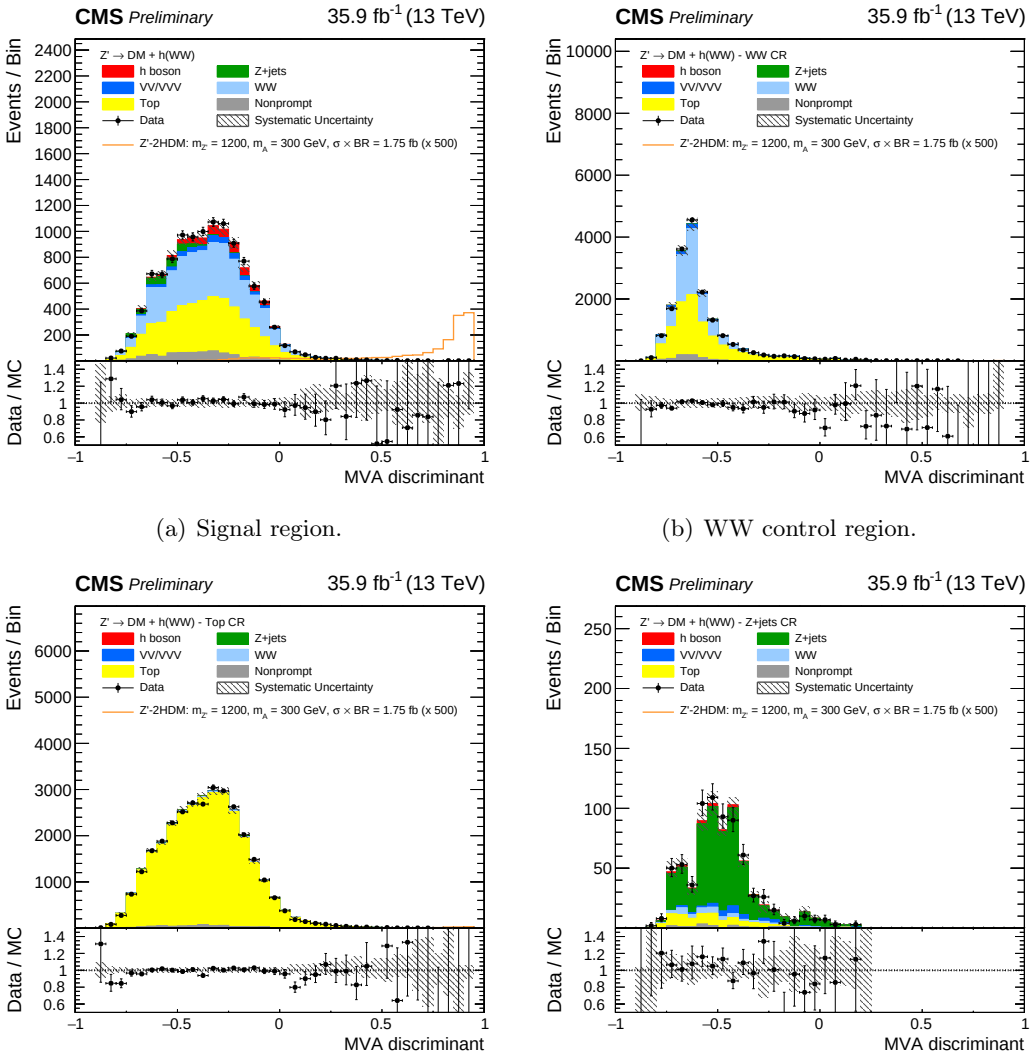


(b)



(c)

Figure 9.2: MVA input variables for Z'-2HDM model. All the signal mass points with $m_A = 300$ GeV, weighted in order to have the same importance, have been used to produce the plots.



(a) Signal region.

(b) WW control region.

(c) Top control region.

(d) Drell-Yan control region.

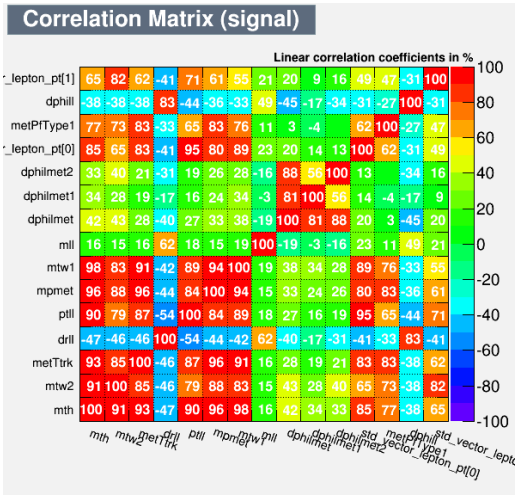
Figure 9.3: BDT distributions for Z'-2HDM model in signal and control regions. Signal events are scaled by a factor 500.

9.1.2 Training for Baryonic Z' Model

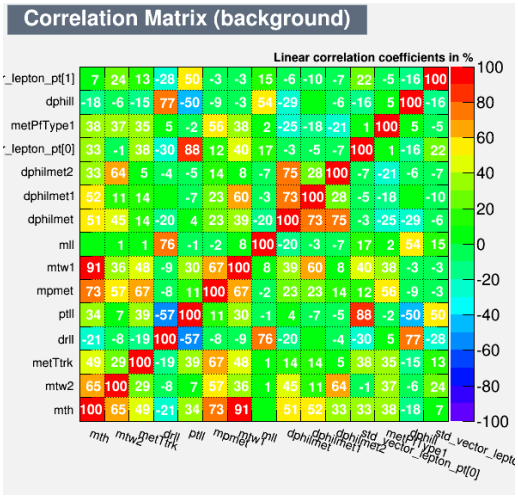
All the signal mass points with $m_\chi = 1$ GeV are considered for the training, since, similarly to the case of the Z'-2HDM model, for heavier masses of the DM particle χ the sensitivity of the analysis rapidly drops. The linear correlations between the variables used for the BDT training are shown in the top part of Figure 9.4. Also for this model, it can be seen that even if some variables are strongly correlated for the signal, the correlation is not so important for the background, so they can add some discriminating power. In the bottom part of Figure 9.4, the ROC curve and the overtraining and performance plots are shown. For the Baryonic Z' model the discrimination is not as strong as in the case of the Z'-2HDM model, due to the different kinematics, which is more similar to the SM one. Nevertheless, good separation between the signal (in blue) and the background (in red) is obtained. Nice agreement between the train and test samples is observed, as expected in case of low level of overtraining. The variables used for the BDT training are shown in Figure 9.5, their ranking, namely the relative contribution to the discrimination power, are presented in Table 9.3, and in Figure 9.6, the BDT output is plotted for data and MC, in the signal region and in the WW, Top, and Drell-Yan control regions, and a good agreement between data and simulation is found.

Rank	Variable	Separation
1	$\min[\Delta\phi(\mathrm{E_T^{miss}}, \ell_1), \Delta\phi(\mathrm{E_T^{miss}}, \ell_2)\ ]$	2.571×10^{-1}
2	$\mathrm{m_T^{\ell\ell \mathrm{E_T^{miss}}}}$	2.418×10^{-1}
3	mpmet	2.147×10^{-1}
4	$\mathrm{m_T^{W_2}}$	2.146×10^{-1}
5	$\mathrm{m_T^{W_1}}$	2.027×10^{-1}
6	$\Delta\phi(\mathrm{E_T^{miss}}, \ell_1)$	1.772×10^{-1}
7	$\Delta\mathrm{R}(\ell, \ell)$	1.738×10^{-1}
8	track. $\mathrm{E_T^{miss}}$	1.678×10^{-1}
9	$\Delta\phi(\mathrm{E_T^{miss}}, \ell_2)$	1.343×10^{-1}
10	$\mathrm{m_{\ell\ell}}$	1.123×10^{-1}
11	PF $\mathrm{E_T^{miss}}$	1.112×10^{-1}
12	$\mathrm{p_T^{\ell\ell}}$	9.406×10^{-2}
13	$\Delta\phi(\ell, \ell)$	9.018×10^{-2}
14	$\mathrm{p_T^{\ell_1}}$	6.109×10^{-2}
15	$\mathrm{p_T^{\ell_2}}$	5.113×10^{-2}

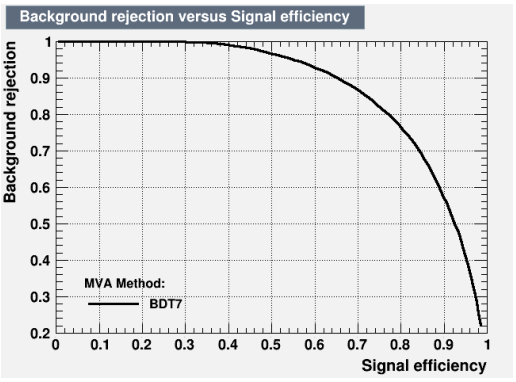
Table 9.3: BDT input variables ranking for Baryonic Z' model.



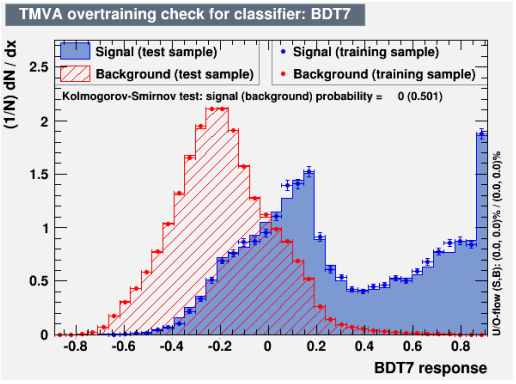
(a) Signal correlation matrix.



(b) Background correlation matrix.

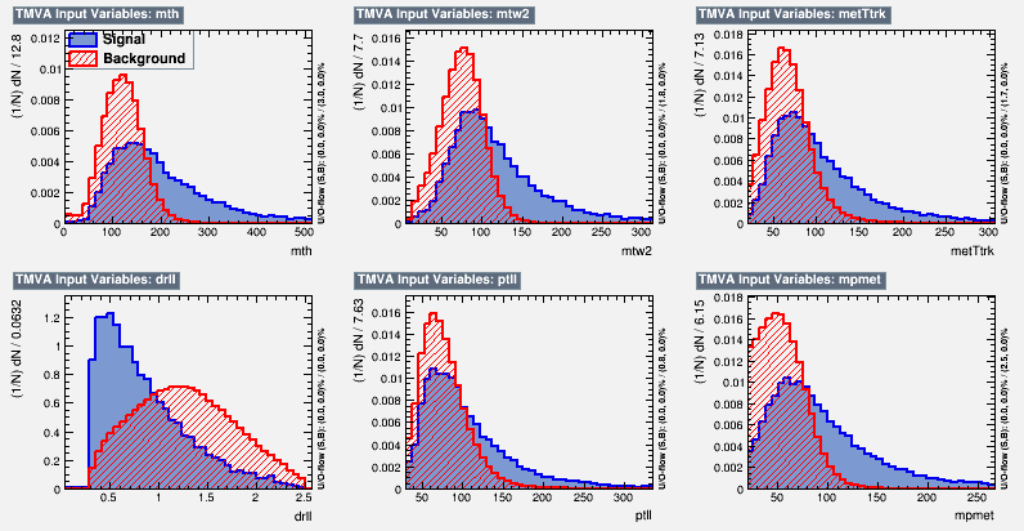


(c) ROC curve.

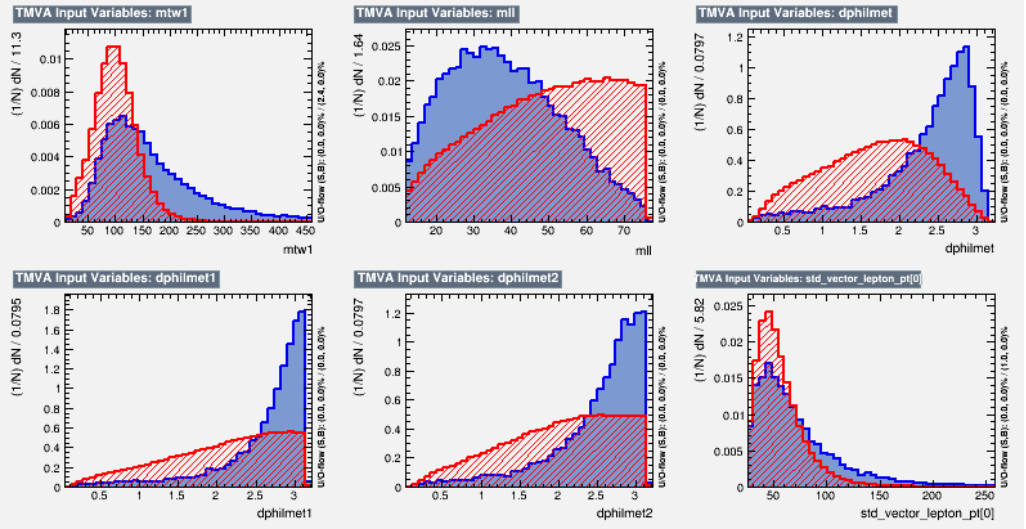


(d) Overtraining plot.

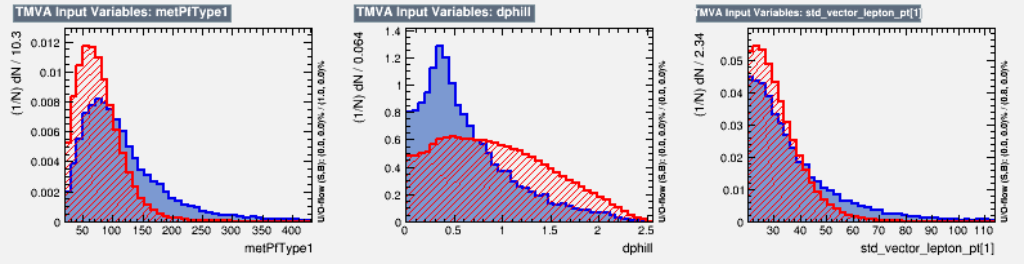
Figure 9.4: Correlation matrix for MVA used in the analysis (top), shown separately for the Baryonic Z' signal (left) and the background (right). The ROC curve (bottom left) and the overtraining and performance plots (bottom right) are also shown.



(a)

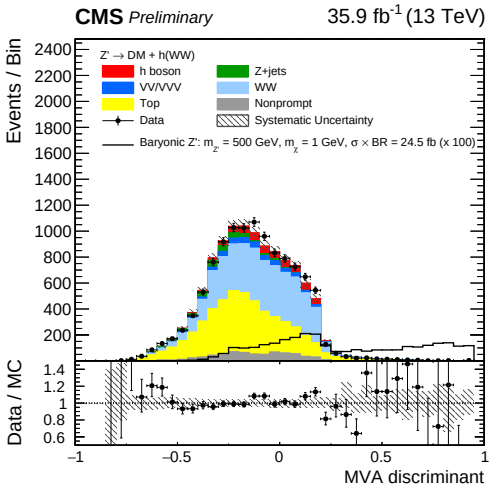


(b)

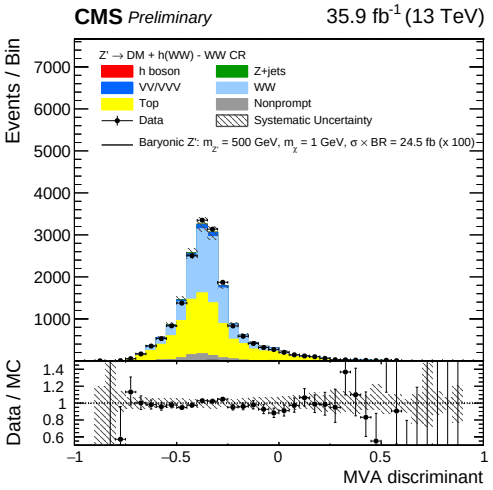


(c)

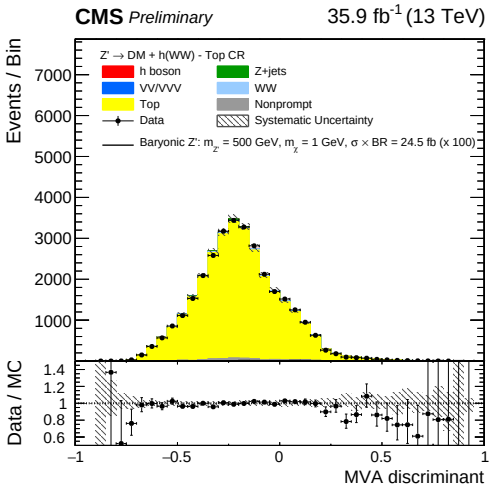
Figure 9.5: MVA input variables for Baryonic Z' model. All the signal mass points with $m_\chi = 1$ GeV, weighted in order to have the same importance, have been used to produce the plots.



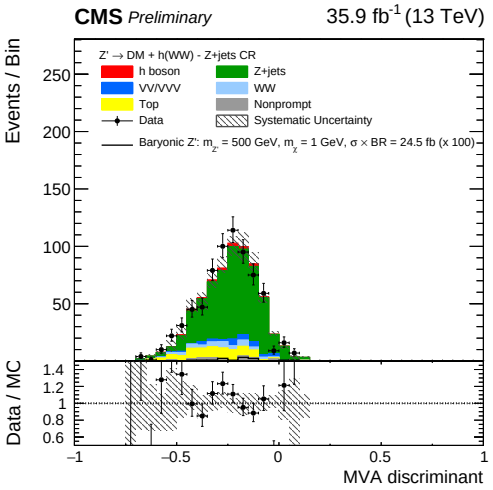
(a) Signal region.



(b) WW control region.



(c) Top control region.



(d) Drell-Yan control region.

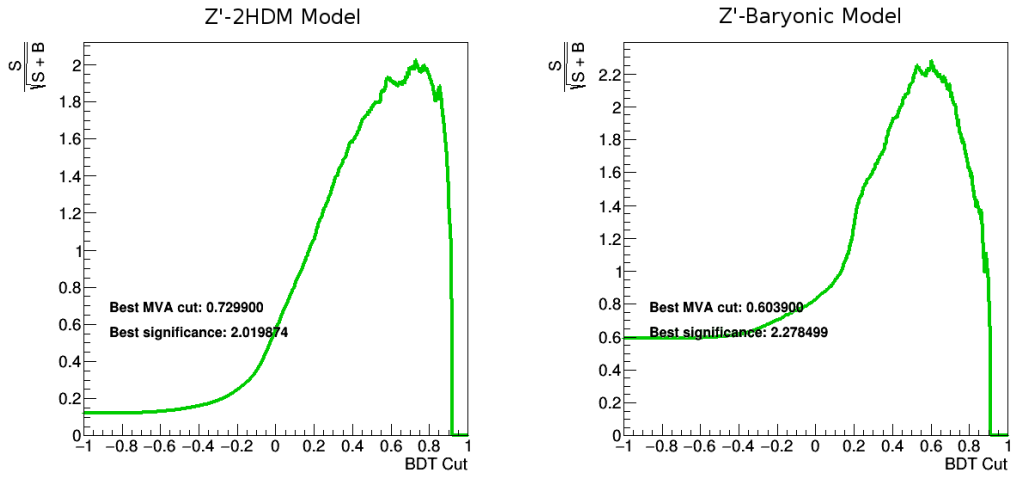
Figure 9.6: BDT distributions for Baryonic Z' model in signal and control regions. Signal events are scaled by a factor 100.

9.2 Shape Analysis

In order to ensure the final likelihood fit to be robust, the binnings applied to the two BDT distributions to extract the signals have been chosen to have a significant number of MC events in each bin. A further step to enhance the sensitivity required to maximize the signal significance $\left(\frac{S}{\sqrt{S+B}}\right)$ in each bin. In particular, for each model, the signal mass point with the highest significance has been taken as reference, such that for each model, only one single binning is defined for all the mass points:

- Z'-2HDM model: $m_{Z'} = 800$ GeV, $m_A = 300$ GeV
- Baryonic Z' model: $m_{Z'} = 100$ GeV, $m_\chi = 1$ GeV

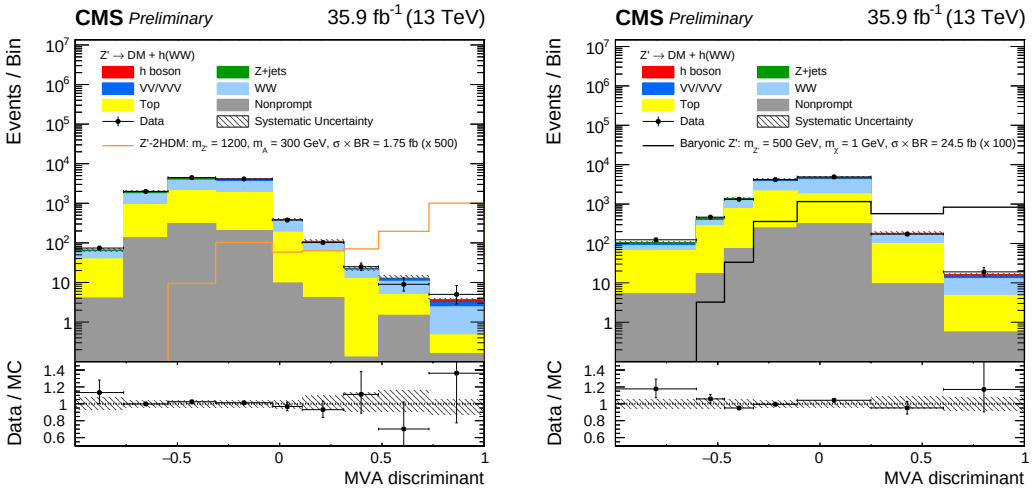
The method starts by defining the last bin to the right of the distribution (closest to +1), which is the most enriched in signal events. To do it, the significance curves (shown in Figure 9.7) are plotted separately for the two BDTs, and considering each one the corresponding signal. The maximum of the distribution is selected as the lower edge of the last bin of the BDT template histogram. For the Z'-2HDM model, this value corresponds to 0.73 (so that the last bin goes from 0.73 to 1), for the Baryonic Z' model it is 0.60 (so that the last bin goes from 0.60 to 1). The method is repeated iteratively until all the bins



(a) Z'-2HDM. (b) Baryonic Z'.

Figure 9.7: Significance distributions used to define the last bin of the BDT template histograms, for the Z'-2HDM model (left) and for the Baryonic Z' model (right).

are defined, as presented in Figure 9.8. Once the binning of the BDT distributions in the signal region has been defined, a maximum likelihood fit is performed to extract the signal. The fit is performed considering the signal region and the Drell-Yan, Top and WW control regions. Since the control regions are used to fix the normalization of specific processes, but not to adjust the BDT shape, each of them is included in the fit as a single bin, and the normalization of the Drell-Yan, Top, and WW processes are let free to float. A different approach has been used for the WZ and Wγ^* processes. In this case, the normalizations and their uncertainties are measured in dedicated control regions, which are not included in the fit (see Section 8.11). The MC predictions of these processes are thus scaled to the measured values in the signal region and in the control regions. More details on the fit are given in the following.



(a) Adaptive boost shape: Z'-2HDM.

(b) Adaptive boost shape: Baryonic Z'.

Figure 9.8: Template histograms for signal extraction based on the BDT output, on the left the one optimized for the Z'-2HDM model, on the right the one optimized for the Baryonic Z' model. The y-axis is plotted in logarithmic scale in order to highlight the bins with less abundant yields. Signal events are scaled by a factor 500 for the Z'-2HDM model and by a factor 100 for the Baryonic Z' model.

9.2.1 Statistical Procedure and Nuisances

The statistical methodology used by the fit has been developed by the ATLAS and CMS collaborations in the context of the LHC Higgs Combination Group [141, 142]. Results presented in this document also make use of asymptotic formulae from [143]. Two quantities are defined to compare the observation in data with the expectation for the analysis:

- The significance, or **p-value**, characterizing the probability of background fluctuations to reproduce an observed excess.
- The **signal strength** ($\sigma/\sigma_{\mathrm{SM}}$) that quantifies the compatibility of the sizes of the possible observed excess with the SM signal expectation.

A description of the statistical formulae defining these quantities is found in [141, 144]. The number of events in each region and in each bin of the discriminant distributions used to extract the signal is modelled as a Poisson random variable, whose mean value is the sum of the contributions from the processes under consideration. Statistical uncertainties from the number of MC simulated events are taken into account. Systematic uncertainties are represented by individual nuisance parameters with log-normal or shape-based distributions. In the first case, the uncertainties affect only the overall normalization of the signal and backgrounds, while in the second case also the shape of the predictions across the distribution of the observables can be moved. The migration of events across bins and correlations between systematic uncertainties in the signal and control regions are taken into account. Three nuisance parameters have been introduced to alter the normalization of the Drell-Yan, Top, and WW processes independently. These nuisance parameters do not contain priors in the likelihood function and can be interpreted as normalization scale factors for each background. Despite the results of the analysis are dominated by the low number of events in the bins of the BDT discriminant distributions with larger

signal significance, systematic uncertainties play an especially important role, since in this analysis no strong mass peak is expected due to the presence of undetected neutrinos in the Higgs final state. In the following of this section, a detailed description of the sources and quantities of systematic uncertainties in this analysis and their effects on the signal and background processes are described. A table presenting the summary of all the systematic uncertainties considered for each process, and the corresponding value assigned is presented in Appendix C.

Experimental Uncertainties

Effects from experimental uncertainties are studied by applying scaling and/or smearing of certain variables of the physics objects, followed by a subsequent recalculation of all the correlated variables. This is done for MC simulation, to account for possible systematic mismeasurements of the data. All experimental sources except luminosity are treated both as normalization and shape uncertainties. For background with a data-driven normalization estimation, the shape uncertainty is considered only. The following experimental systematic sources have been taken into account:

- **Luminosity:** The uncertainty determined by the CMS luminosity monitoring [145] is 2.5% for the full 2016 13 TeV data.

- **Lepton trigger systematics:** Lepton trigger systematics are of the order of 2%. These uncertainties are computed by varying the tag selection, and hence the background contribution, as well as the Z window, in the Tag and Probe method.

- **Lepton reconstruction and identification efficiency:** The lepton reconstruction and identification efficiencies are measured with the Tag and Probe method in data and MC. To correct for the difference in the lepton identification efficiencies between data and MC, data/MC scale factors dependent on p_T and η are applied to the MC. The uncertainties on the scale factors vary between ∼0.5-5% for electrons and ∼1-7% for muons.

- **Lepton energy scale and resolution:** These uncertainties arise due to different detector effects and are p_T and η-dependent. Uncertainties on both the scale and resolution individually amount to ∼0.1-0.5% for electrons and ∼0.5-1.5% for muons. The numbers for the electrons are taken from [146]. For muons the procedure described below has been followed:

  - **Muon momentum scale:** The dimuon invariant mass distribution is fitted using Gaussian distribution and the shift in the Z peak is observed for data and simulated Drell-Yan sample. The relative shift in the Z peak for data and simulated sample is measured to be of the order of 0.04%. This effect is also studied as a function of muon η and ϕ to avoid any bias. A discrepancy in the residual SF comparing data and simulation of about ∼0.5% in the central region (muon $|\eta| < 2.2$) and of the order of ∼1.5% for the forward region (muon $|\eta| > 2.2$) has been observed. The final systematic uncertainty for muons is 1.3% [147].
  - **Muon resolution uncertainty:** Comparing the shape of the dimuon invariant mass distribution for data and simulated Drell-Yan sample, the RMS value of the peak matches well within uncertainty, hence this amounts to very negligible effect. The analysis is not sensitive to a resolution effect.

The final lepton scale nuisance includes as well the effect of the lepton momentum scale on the E_T^{miss}, following the recommendation of the JET-MET POG [148].

- **Jet energy scale (JES) uncertainties:** The estimation of this uncertainty follows the official recommendation of the JET-MET POG [148]. The energy of the jets is moved up and down by a factor which depends on η and p_T of the jet, and the variation of the selection efficiency is computed. The effect of moving the energy scale of the jets on the E_T^{miss} is included in this nuisance, following the recommendation given by the JET-MET POG. JES uncertainty varies between ∼1-11%.

- **E_T^{miss} modelling:** The E_T^{miss} measurement is affected by the possible mismeasurement of individual particles addressed above, as well as the additional contributions from the pile-up interactions. Since the effect of moving jet and lepton energy scale on the E_T^{miss} is already taken into account in the lepton momentum scale and jet energy scale nuisances, here only the unclustered PF candidates scale uncertainty and its effect on the E_T^{miss} is considered.

- **b-jet misidentification modelling:** The uncertainties on the selection of b-jets when requiring b-veto, or non-b-jets in the b-tagged region is taken into account by looking at the b-jet misidentification efficiency. The differences in data and MC of the b-jet misidentification efficiency are corrected by applying p_T, η and jet flavour-dependent scale factors to the MC events. The uncertainties on these scale factors need to be taken into account and are of the order of few percents. In particular, this systematic uncertainty is anticorrelated between the top control region and the signal region. The corrections and the uncertainties are provided down to 30 (20) GeV for b (light) jets. Where not available, for example for b-jets in the 20-30 GeV p_T range, the BTV recommendation [131] to apply the same scale factor, but doubling the uncertainty has been followed. The uncertainty on the signal efficiency is at the level of 1%.

Theoretical Uncertainties

The description of the main theoretical uncertainties affecting the analysis follows:

- **PDF and higher-order corrections (renormalization and factorization scales):** The choice of a specific set of PDF and the missing knowledge on higher-order corrections, evaluated by means of scale variation, directly affect the cross-section, as well as the acceptance of a simulated process. The uncertainty due to PDF is computed following the PDF4LHC recommendations for LHC Run 2 [149]. In case of the Higgs background, the PDF and α_s uncertainties are further split between the cross-section normalization uncertainties, computed by the LHC Higgs cross-section working group [150], and the effect of varying the PDFs and α_s within their uncertainties on the acceptance [149]. The PDFs and α_s cross-section normalization uncertainties are $^{+7.4\%}_{-7.9\%}$ and $^{+7.1\%}_{-6.0\%}$ for gluon fusion Higgs production, ±0.7% and ±3.2% for VBF Higgs production, ±1.3% and ±0.9% for ZH production, ±1.7% and ±0.9% for WH production mechanism. The effect of varying the choice of the PDF and α_s scale on the acceptance is less than 1% for gluon fusion and 1% for VBF, ZH and WH Higgs production mechanisms. The uncertainties on the yields from missing higher-order corrections are evaluated by independently varying up and down the factorization and renormalization scales by a factor of two. The uncertainty on the acceptance has been tested not to change the shape of the final distributions, but only affecting the

selection efficiency. It has then been estimated for each MC based sample and it is modelled as a log-normal prior. For each background control region, the effect of the PDF and QCD scales has been added as an extra nuisance. In this way, any possible mismodelling of the simulation (even if small) due to these theoretical systematics is taken into account.

- **Underlying event and parton shower modelling:** The underlying event (UE) and parton shower (PS) modelling uncertainties are estimated by comparing samples interfaced with different parton showers (Pythia vs Herwig) and UE tunes. The effect results to be about 3% for the WW and $h \to W^+W^-$ processes, for both the PS and UE.

- **Single top tW and $t\bar{t}$ ratio:** Since the single top and $t\bar{t}$ processes are normalized together in the same control region, and the fraction of each process can be different in the control region and in the signal region, the uncertainty on the ratio of the cross-sections of the two processes has to be taken into account. The ratio between the single top and top pair cross-sections is varied by the uncertainty on the ratio between their cross-sections, calculated considering scale variations, PDF variation and mass uncertainty, following the recipe in [151]. From [152] and [153] the $t\bar{t}$ and single top cross sections are respectively 831.76 pb and 71.70 pb, with uncertainties of 6% and 5%, respectively. The uncertainty on the ratio is found to be 8%.

- **ggWW:** The k-factor applied to the LO $gg \to WW$ cross-section is 1.4, with an uncertainty of 15% (see [154]).

- **qqWW modelling:** The p_T^{WW} spectrum of the $qq \to WW$ process has been reweighed to NNLL+NNLO resummed calculation, as explained in Section 8.8. The uncertainty related to missing higher-orders is modelled varying the factorization and renormalization scale by a factor 2, and by varying the resummation scale as well. This uncertainty is then treated as shape uncertainty, allowing shaping of the final BDT output distribution. The acceptance of the process varies between 1% and 5% due to these effects.

Background Normalization Uncertainties

One of the most important sources of systematic uncertainty is the normalization of the backgrounds that are estimated on data control samples whenever possible. The signal extraction is performed subtracting the estimated backgrounds to the event counts in data. The amount of uncertainty depends on the considered background.

- **Non-Prompt Background:** Normalization and kinematic shapes are derived from a data control region and both normalization and shape systematic uncertainties are considered. The uncertainty is estimated to be of the order of 15%, including both the statistical part of the fake-rate computation and the variation of the jet-E_T cut used in defining the control region sample for computing the fake-rate. These uncertainties do not cover the discrepancies observed in the same-sign validation regions. An additional flat 30% uncertainty is considered to cover such discrepancies.

- **WW background:** The kinematic shape of this background is predicted by simulation, and a control region is defined as a $m_{\ell\ell}$ sideband with respect to the signal region to constrain its contribution in the signal region. The uncertainty related to the normalization of this background is about 11%, and is given by the result of the

simultaneous fit to the signal region and all the control regions, for which the WW normalization is a free-to-float parameter. This uncertainty value includes all the theoretical and experimental nuisances affecting the process.

- **$t\bar{t}$ and tW backgrounds:** Top events are estimated in a dedicated b-tagged control region. The top background enriched control region is defined as additional categories in the fit while the kinematic shape is taken from the simulation, corrected for the b-tagging discriminant scale factors (as explained in Section 8.9). The top normalization is correlated between the top control region and the signal region. The uncertainty assigned to the normalization of this background is about 2%, and takes into account all the theoretical and experimental nuisances affecting the process. An additional 1% is given to this background normalization, to take into account the uncertainty in the extrapolation between the control region and the signal region. This is necessary since the control region is made of just one bin, so that to take into account effects of shape mismodelling (even if small) given by PDF and QCD scale variations, the change in yields in the control region (b-tagged) due to the migration of events to the signal region (b-vetoed) has been measured for the $t\bar{t}$ MC sample and considered as uncertainty.

- **Drell-Yan background:** The kinematic shape of this background is predicted by simulation, and a control region enriched in this background (see Section 8.10) is included in the fit to further constrain its contribution in the signal region. The uncertainty assigned to the normalization of this background is about 11% and takes into account all the theoretical and experimental nuisances affecting the process. An additional 2% is given to this background normalization, to take into account the uncertainty in the extrapolation between the control region and the signal region. This is necessary since the control region is made of just one bin, so that to take into account effects of shape mismodelling (even if small) given by PDF and QCD scale variations, the change in yields in the control region (low $m_T^{\ell\ell E_T^{miss}}$) due to migration of events to the signal region (high $m_T^{\ell\ell E_T^{miss}}$) has been measured for the Drell-Yan MC sample and considered as uncertainty.

- **WZ and $W\gamma^{(*)}$ backgrounds:** The kinematic shape of these backgrounds is predicted by simulation, normalized to its data-driven estimate, and constrained within the respective uncertainty, which is 16% for WZ and 26% for $W\gamma^*$ (see Section 8.11).

- **Other minor backgrounds:** The kinematic shapes of these backgrounds are predicted by simulation and normalized to their theoretical predictions.

Chapter 10

Results and Interpretation

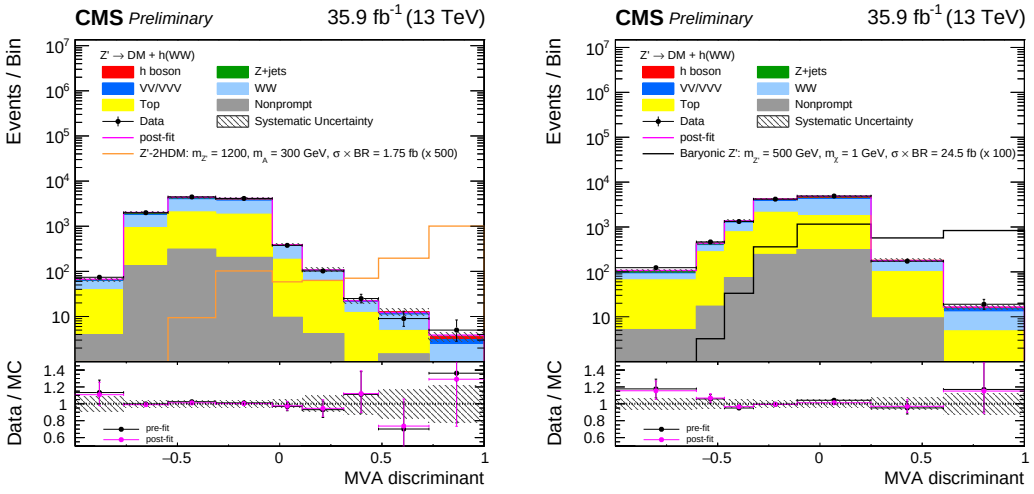
È una questione di qualità?
O una formalità?

CCCP, *Io sto bene*

The results of the fit are presented in this chapter as signal and background yields, with their relative uncertainties. Since no significant discrepancies with respect to the Standard Model predictions are observed in data, they are interpreted as upper limits on the production cross section of Dark Matter in proton-proton collisions at a centre of mass energy of 13 TeV. Fixing the signal cross sections to the theoretical predictions for each mass point, it is thus possible to interpret the results in terms of mass points excluded by the observations. A limited number of signal mass points has been simulated in detail, so that an interpolation technique, which makes use of generator-level information, has been exploited to obtain smooth exclusion maps in the ($m_{Z'}$,m_A) plane for the Z'-2HDM model, and in the ($m_{Z'}$,m_χ) plane for the Z'-baryonic model.

10.1 Z'-2HDM Model Results

The results of the fits to the BDT discriminants trained for the Z'-2HDM model are presented in this Section. A visual comparison between the pre-fit and post-fit distributions is shown in the left part of Figure 10.1, while in Table 10.1 and Table 10.2 a quantitative way of displaying the results is presented. In particular, the yields in the last and in the penultimate bins of the BDT output discriminant (which are expected to drive the significance of the analysis) and the corresponding uncertainties before and after the fit are listed. The results show how the fit does not significantly change the yields obtained from the MC prediction, which is a sign of the robustness of the simulation. On the other hand, particularly for the main backgrounds (WW and Top), the post-fit uncertainties are reduced with respect to the pre-fit values, thanks to the introduction of the control regions. Several checks have been performed to ensure the goodness of the fit. In Figure 10.2, the so-called *impact plot* for the Z'-2HDM model fit is presented. Here both the *pulls* and the *impacts* of the 30 most important sources of uncertainty affecting the results are shown. A source of uncertainty (also called nuisance) represents a parameter of the fit. Such parameter has, in general, a central value and a gaussian uncertainty. Let's consider for example the lepton momentum scale: the central value is given by the scale actually used in the analysis and



(a) Adaptive boost shape: Z'-2HDM.

(b) Adaptive boost shape: Z' baryonic.

Figure 10.1: Pre-fit and post-fit distributions for the BDT output, on the left the one optimized for the Z'-2HDM model, on the right the one optimized for the Z' baryonic model. The red line in the upper part of the plots indicates the sum of all the backgrounds after the fit. In the ratio plot, black points refer to pre-fit MC, while red points to post-fit MC. Signal events are scaled by a factor 500 for the Z'-2HDM model and by a factor 100 for the Z'-Baryonic model.

the uncertainty, provided by CMS, affects the shape of the BDT spectrum and the efficiency of each process in a different way. The fit adjusts the MC predictions to data and can move the central value of each nuisance parameter (a nuisance parameter with large uncertainty can be moved more than a nuisance parameter with small uncertainty) and reduce, in some cases the relative uncertainty, for example in a region with large amount of data, with a better data/MC agreement than the one expected by the pre-fit uncertainty. In the central column in Figure 10.2, the pulls for the Z'-2HDM analysis are shown. The displacement of the central value with respect to the origin represents how much the nuisance parameter has been moved by the fit, in unities of pre-fit uncertainty. The amplitude of the error bar represents the ratio between the post-fit and the pre-fit uncertainty related to the nuisance parameter. As an example, let's consider the uncertainty 6 (CMS_scale_m), corresponding to the muon momentum scale: the central value is slightly shifted to the left, which means that the fit preferred a muon momentum scale slightly lower than the nominal one, more specifically about 0.2σ smaller than the pre-fit value. Additionally, the error bars do not reach the -1 or the +1 vertical columns: this means that the uncertainty on this parameter has been *constrained* by the fit, such that the post-fit uncertainty is about 0.6 times the pre-fit uncertainty. In general, a fit which converged well presents uncertainties slightly constrained or not constrained, but never post-fit uncertainties larger than pre-fit uncertainties. Regarding the central values, shifts of more than 1σ are expected to be quite rare. The fit presented here satisfies both these sanity-check requirements so that the results obtained are considered as reliable. An exception to what explained above is represented by the nuisance parameters left unconstrained before the fit (in this case, the normalization parameters WWnorm, Topnorm and DYttnorm). For these nuisances, no central value is assigned, and the pre-fit uncertainty can be considered infinite. For this reason, instead of the pulls with respect to the pre-fit values, the post-fit central values

and uncertainties are directly given numerically and are interpreted as normalization scale factors for the process affected and relative uncertainty. In the three cases considered, the scale factors are compatible with 1 within one standard deviation. In the right column of the same figure, the *impacts* of the individual nuisance parameters on the signal strength are shown, with the uncertainties with the largest impact on the top. The impact of a nuisance parameter is measured by repeating the fit with all the nuisances set to their nominal values, with the exception of the nuisance considered, which is moved by $\pm 1\sigma$. The relative change in the signal strength when the parameter is moved up by one standard deviation is shown in red, while the change in the result obtained by moving the parameter down by 1σ is shown in blue. Considering the second nuisance (CMS_scale_j), corresponding to the jet energy scale, the impact plot shows that decreasing the nuisance by 1σ enhances the signal strength by 4% while increasing the jet energy scale by 1σ reduces the signal strength by 7%. The nuisance with the largest impact is the uncertainty related to the unclustered energy, which moves the value of the signal strength by 20%. On the other hand, the global uncertainty on the signal strength is much larger and is mainly given by the low number of data events in the bins of the BDT distribution with the largest signal significance. No evidence of new physics have been found in data, so that the results are interpreted in terms of upper limits on the DM production cross section, or in terms of mass points ruled-out by the analysis, assuming the theoretical cross section predicted by the model, as shown in Figure 10.3. The plot on the top shows the results in terms of limits on the signal strength, namely the ratio between the measured cross section and the theoretical cross section. The limits are presented as a function of the Z' mediator mass and fixing the pseudoscalar mass m_A to 300 GeV. The results obtained considering only the mass points for which fully simulated MC samples (see Table 7.2) have been produced are presented in this plot. In the bottom plot of the same figure, the results are presented as a function of the Z' mediator mass (x-axis) versus the A pseudoscalar mass (y-axis). On the z-axis, the signal strength is plotted, such that mass points with a signal strength lower than 1 are excluded. In particular, the analysis is able to rule out mass points with $m_{Z'}$ between 740 GeV and 800 GeV, and m_A lower than 320 GeV. To present this smooth map, a larger set of samples has been produced at generator level. The ratio of the generator-level Higgs boson p_T between fully simulated samples and generator-level samples is used to weight the fully simulated samples to estimate the kinematic distributions of the mass points produced at generator-level. This method assumes that reweighting the Higgs boson p_T is sufficient to reproduce in a sensible way the kinematics of the additional mass points, and was validated by applying the same procedure among fully simulated samples (see Appendix D). The results take into account the correct cross sections also for the MC samples produced at generator-level only.

	Pre-Fit		Post-Fit	
Process	Yields	Uncertainty	Yields	Uncertainty
WW	1.88	0.55	2.01	0.57
ggWW	0.04	0.01	0.04	0.01
top	0.31	0.06	0.30	0.05
Fake	0.16	0.24	0.23	0.25
DY	0.00	0.00	0.00	0.00
WH_hww	0.12	0.01	0.12	0.01
ZH_hww	0.23	0.02	0.23	0.01
bbH_hww	0.00	0.00	0.00	0.00
ggH_hww	0.00	0.00	0.00	0.00
ggZH_hww	0.24	0.10	0.24	0.08
qqH_hww	0.00	0.00	0.00	0.00
H_htt	0.01	0.00	0.01	0.00
WZgS_H	0.25	0.04	0.25	0.04
WZgS_L	0.19	0.05	0.19	0.05
VVV	0.10	0.03	0.10	0.02
VZ	0.15	0.01	0.15	0.01
Vg	0.00	0.00	0.00	0.00
total_background	3.68	0.81	3.87	0.83
monoH_1200_300	2.01	0.06	1.03	1.73
Data	5	2.84	5	2.84

Table 10.1: Pre-fit and post-fit yields and corresponding uncertainties in the last bin of the discriminant distribution for the Z'-2HDM model.

	Pre-Fit		Post-Fit	
Process	Yields	Uncertainty	Yields	Uncertainty
WW	7.01	1.34	6.85	1.26
ggWW	0.64	0.12	0.63	0.10
top	3.74	0.56	3.51	0.42
Fake	1.63	0.69	1.63	0.85
DY	0.00	0.14	0.02	0.09
WH_hww	0.20	0.03	0.20	0.02
ZH_hww	0.48	0.02	0.48	0.02
bbH_hww	0.00	0.00	0.00	0.00
ggH_hww	0.00	0.05	0.00	0.02
ggZH_hww	0.59	0.17	0.60	0.14
qqH_hww	0.02	0.01	0.02	0.01
H_htt	0.04	0.00	0.04	0.00
WZgS_H	0.56	0.07	0.56	0.07
WZgS_L	0.39	0.07	0.39	0.07
VVV	0.48	0.05	0.48	0.04
VZ	0.34	0.02	0.34	0.02
Vg	0.37	0.09	0.34	0.09
total_background	16.49	2.31	16.09	2.05
monoH_1200_300	2.40	0.06	1.24	1.77
Data	9	3.57	9	3.57

Table 10.2: Pre-fit and post-fit yields and corresponding uncertainties in the penultimate bin of the discriminant distribution for the Z'-2HDM model.

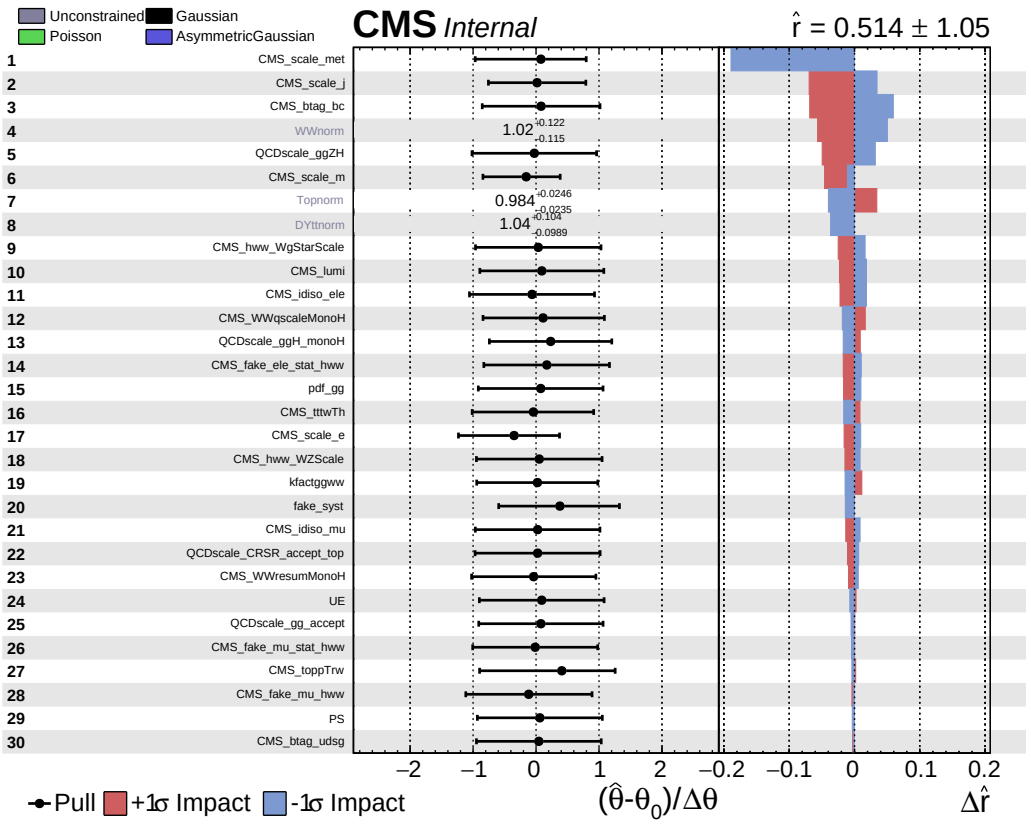
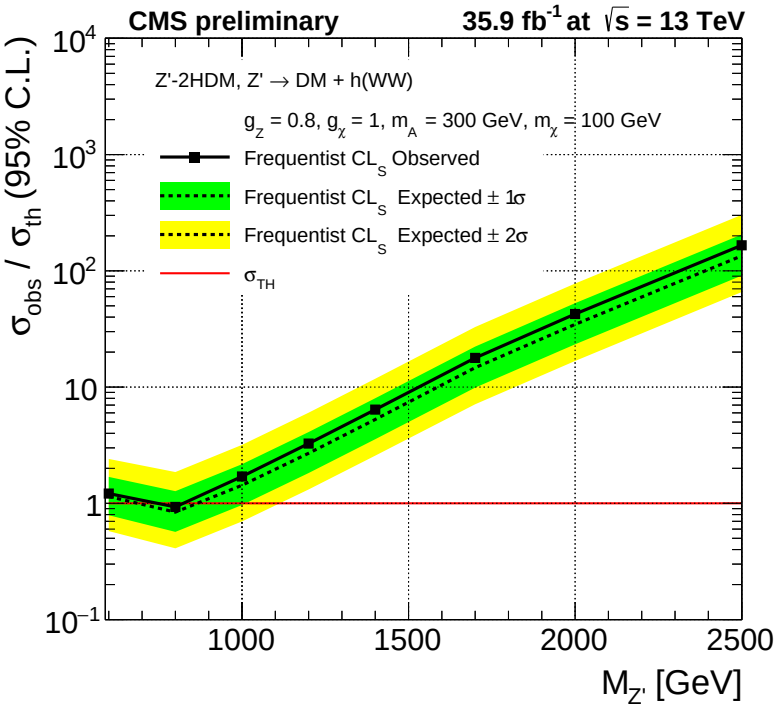
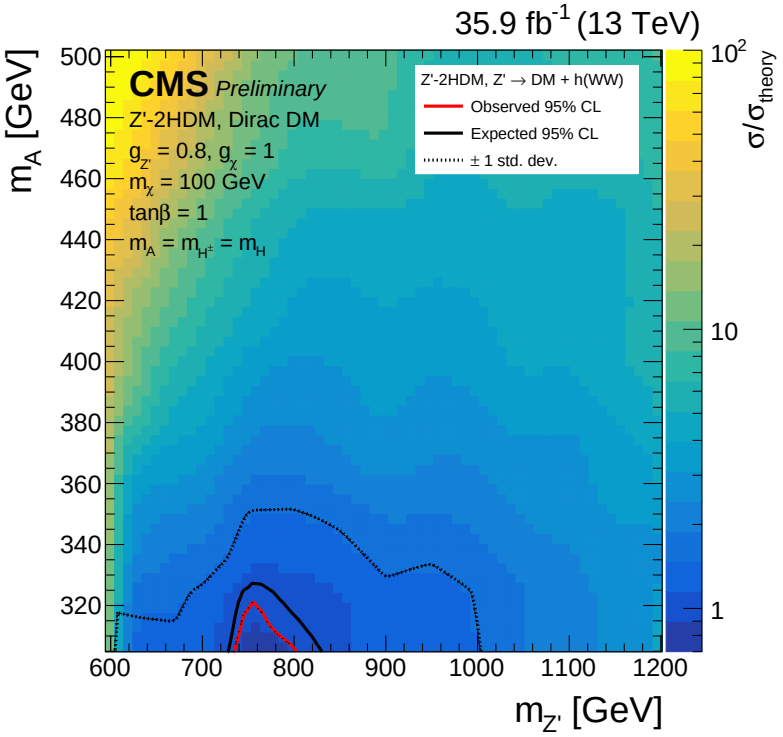


Figure 10.2: Impact plot for the fit of the $m_{Z'} = 1200$ GeV, $m_A = 300$ GeV mass point in the Z'-2HDM model. The 30 uncertainties with the largest impact on the final results are shown. All the uncertainties have a gaussian prior, with the exception of the backgrounds normalizations, which are unconstrained. In the central column, the *pulls* of the uncertainties, which are considered as parameters of the fit, are shown for gaussian uncertainties, while the normalization scale factors and their uncertainties are shown for the unconstrained nuisances. In the right column, the impact of each uncertainty on the final result after the fit is illustrated.



(a) $m_{Z'}$ limits.



(b) $m_{Z'}$ vs m_A limits.

Figure 10.3: Limits on the Z'-2HDM model. On the top, limits on the signal strength are shown as a function of the Z' mediator. On the bottom, the expected (black) and the observed (red) mass points excluded by the analysis are shown in the ($m_{Z'}$,m_A) plane.

10.2 Z'-Baryonic Model Results

The results of the fit to the BDT discriminant trained for the Z'-Baryonic model are presented in this section. Similarly to the Z'-2HDM model case, a visual comparison between the pre-fit and post-fit distributions is shown in the left part of Figure 10.1, while in Table 10.3 and Table 10.4 the yields in the last bin and in the penultimate bin of the BDT output discriminant and the corresponding uncertainties before and after the fit are listed. The results show how the fit does not significantly change the yields obtained from the MC prediction, which is a sign of the robustness of the simulation. On the other hand, particularly the Top process, the post-fit uncertainty is significantly reduced with respect to the pre-fit value, thanks to the introduction of the control region. Also for the Z'-Baryonic interpretation, sanity checks for the fit have been performed, and are presented in Figure 10.4 as pulls and nuisances impacts on the signal strength. In this case, the main sources of systematic uncertainty are the lepton momentum scales (up to 9% on the final results), which in any case give a contribution much lower than the statistical uncertainty. No significance discrepancies between data and the Standard Model predictions have been found, so that the results are interpreted in terms of upper limits on the DM production cross section, or in terms of mass points ruled-out by the analysis, assuming the theoretical cross section predicted by the model, as shown in Figure 10.5. The plot on the top of the figure shows the results in terms of limits on the signal strength. The limits are presented as a function of the Z' mediator mass and fixing the dark matter particle mass m_χ to 1 GeV. The results obtained considering only the mass points for which fully simulated MC samples (see Table 7.2) have been produced are presented in this plot. In the bottom plot of the same figure, the results are presented as a function of the Z' mediator mass (x-axis) versus the dark matter particle mass (y-axis). On the z-axis, the signal strength is plotted, such that mass points with a signal strength lower than 1 are excluded. In this case, the analysis is not able to rule out ant mass points: a slight excess in data, even if not incompatible with the SM makes it impossible to exclude any mass points for this model. Similarly to the Z'-2HDM case, a larger set of samples has been produced at generator level, and the ratio of the generator-level Higgs boson p_T between fully simulated samples and generator-level samples has been used to weight the fully simulated samples to estimate the kinematic distributions of the mass points produced at generator-level, in order to obtain a smooth map for the results.

	Pre-Fit		Post-Fit	
Process	Yields	Uncertainty	Yields	Uncertainty
WW	7.01	1.04	7.55	1.14
ggWW	0.83	0.25	0.83	0.16
top	4.23	0.45	3.99	0.30
Fake	0.56	0.42	0.68	0.47
DY	0.00	0.00	0.00	0.00
WH_hww	0.24	0.03	0.24	0.02
ZH_hww	0.51	0.04	0.51	0.03
bbH_hww	0.00	0.00	0.00	0.00
ggH_hww	0.00	0.05	0.00	0.03
ggZH_hww	0.61	0.22	0.63	0.23
qqH_hww	0.02	0.00	0.01	0.00
H_htt	0.03	0.00	0.03	0.00
WZgS_H	0.58	0.12	0.57	0.11
WZgS_L	0.44	0.13	0.46	0.12
VVV	0.46	0.04	0.45	0.03
VZ	0.34	0.02	0.34	0.02
Vg	0.37	0.34	0.21	0.21
total_background	16.23	1.87	16.50	1.77
monoH_ZB_500_1_	8.34	0.29	2.79	3.37
Data	19	4.91	19	4.91

Table 10.3: Pre-fit and post-fit yields and corresponding uncertainties in the last bin of the discriminant distribution for the Z'-Baryonic model.

	Pre-Fit		Post-Fit	
Process	Yields	Uncertainty	Yields	Uncertainty
WW	59.45	3.82	62.49	5.08
ggWW	18.40	2.91	18.65	2.40
top	94.76	7.99	88.04	4.59
Fake	9.91	3.14	10.35	3.23
DY	0.22	0.70	0.24	0.57
WH_hww	0.78	0.07	0.78	0.05
ZH_hww	0.98	0.05	0.98	0.05
bbH_hww	0.01	0.00	0.01	0.00
ggH_hww	1.55	0.52	1.51	0.32
ggZH_hww	1.09	0.28	1.13	0.29
qqH_hww	0.22	0.05	0.20	0.04
H_htt	0.09	0.01	0.09	0.01
WZgS_H	3.00	0.45	2.89	0.45
WZgS_L	1.12	0.24	1.20	0.21
VVV	1.10	0.12	1.10	0.09
VZ	0.84	0.05	0.84	0.04
Vg	4.48	1.63	4.27	1.31
total_background	198.00	13.20	194.77	7.68252
monoH_ZB_500_1_	14.05	0.36	4.72	4.11
Data	173	13.67	173	13.67

Table 10.4: Pre-fit and post-fit yields and corresponding uncertainties in the penultimate bin of the discriminant distribution for the Z'-Baryonic model.

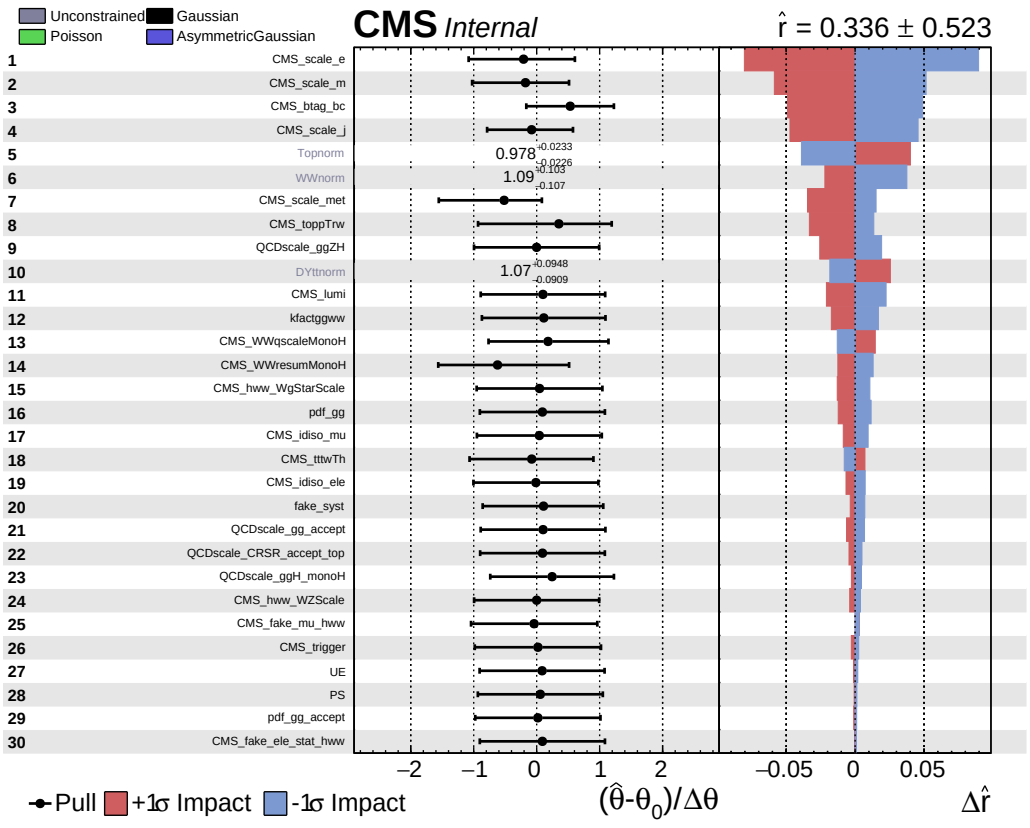
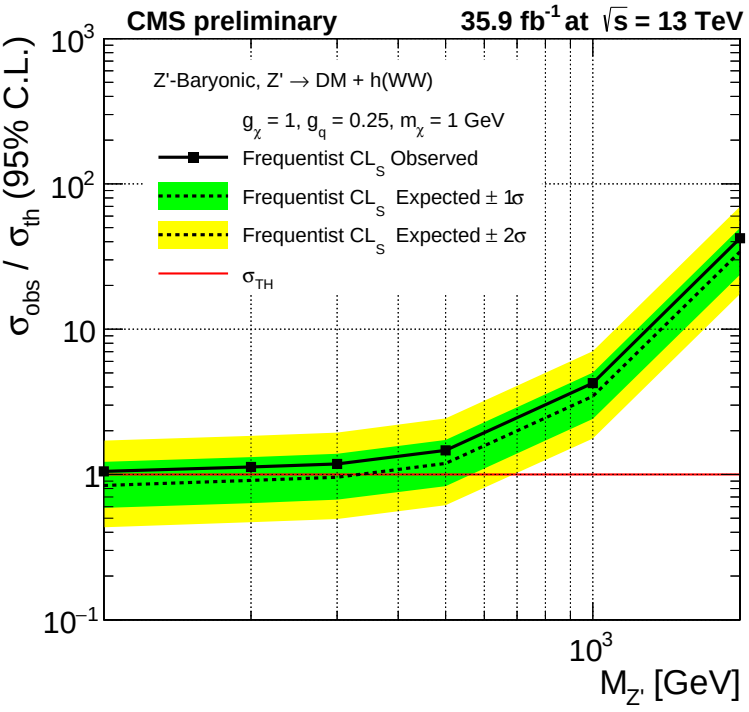
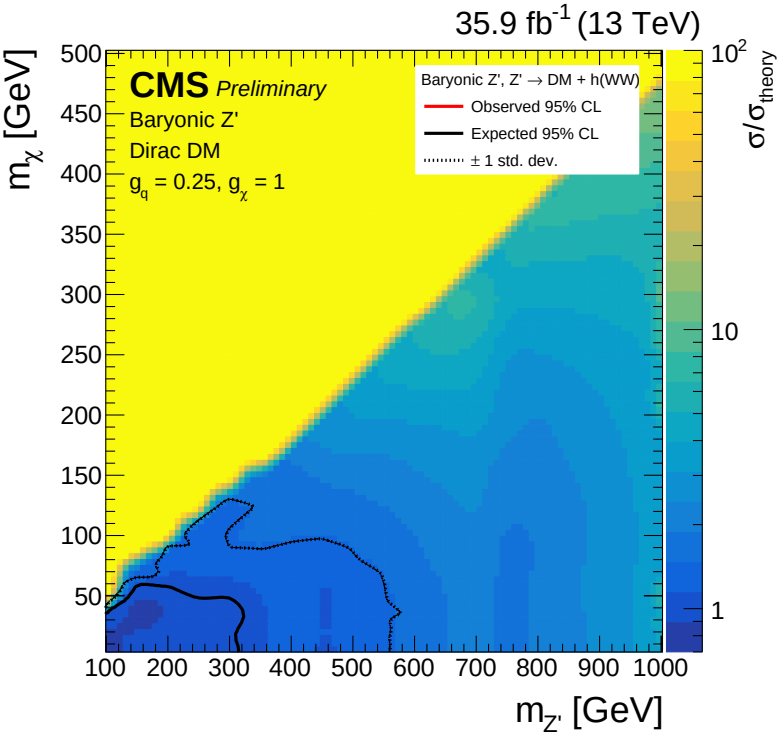


Figure 10.4: Impact plot for the fit of the $m_{Z'} = 500$ GeV, $m_{\chi} = 1$ GeV mass point in the Z'-Baryonic model. The 30 uncertainties with the largest impact on the final results are shown. All the uncertainties have a gaussian prior, with the exception of the backgrounds normalizations, which are unconstrained. In the central column, the *pulls* of the uncertainties, which are considered as parameters of the fit, are shown for gaussian uncertainties, while the normalization scale factors and their uncertainties are shown for the unconstrained nuisances. In the right column, the impact of each uncertainty on the final result after the fit is illustrated.



(a) $m_{Z'}$ limits.



(b) $m_{Z'}$ vs m_{χ} limits.

Figure 10.5: Limits on the Z'-Baryonic model. On the top, limits on the signal strength are shown as a function of the Z' mediator. On the bottom, the expected (black) mass points excluded by the analysis are shown in the ($m_{Z'}$,m_{χ}) plane. The observation does not allow to exclude any mass points.

Chapter 11

Conclusions

E la morale
di questa storia
è che la
non è così brutta
come la si dipinge.

Elio e le storie tese, *Cateto*

In this document, the search for dark matter produced in association with a Higgs boson through proton-proton collisions at the LHC at a centre of mass energy of 13 TeV has been presented. The results obtained are based on data collected by the CMS experiment during 2016, for a total integrated luminosity of 35.9 fb^{-1}. The Higgs boson decay channel selected is to a pair of W bosons, both decaying to a lepton (an electron or a muon) and the corresponding neutrino, such that the final state inspected presents two well identified and isolated leptons of different flavour and opposite charge, and a significant amount of E_T^{miss}. Two simplified models, introducing additional particles to the Standard Model of particle physics, have been used as benchmarks. The first one, denominated Z'-2HDM model, introduces a new Z' vector mediator, which can be produced by quark-antiquark annihilation, and then decays to a pseudoscalar mediator A and a Higgs boson. The pseudoscalar A then decays to a dark matter particles pair. In the second model, called Baryonic-Z' model, a baryonic mediator Z' is produced by quark-antiquark annihilation, irradiates a Higgs boson, and finally decays to a pair of dark matter particles. The search for dark matter produced in association with a Higgs boson, also called mono-Higgs search, is usually based on selecting events where the presence of the Higgs boson is tagged through invariant mass requirements, and the presence of dark matter is expected to produce a large amount of missing energy in the transverse plane. The presence of the neutrinos spoils both the invariant mass distribution and the E_T^{miss} spectrum, and this standard approach cannot be directly followed. Multivariate analysis techniques, specifically boosted decision trees, have been employed to recover the sensitivity of the analysis to those models also when the fully leptonic WW final state is considered. The selected samples of mono-Higgs(WW) candidate events are compared to the estimation of the background processes, simulated with Monte Carlo techniques or estimated from experimental data. No statistically significant discrepancies with respect to the Standard Model predictions have been found observing data so that the results are interpreted as upper limits on the production cross-section of dark matter through proton-proton collisions. Since the models inspected provide predictions for the production cross sections, depending on the mass point, these

limits are also displayed as maps showing the mass points ruled out by the observations. For the Z'-2HDM model, the analysis is able to rule out mass points with $m_{Z'}$ between 740 GeV and 800 GeV, and m_A lower than 320 GeV, while for the Z'-Baryonic model no mass points have been excluded. The results obtained in this search have been combined with those produced by other mono-Higgs analyses (bb, $\gamma\gamma$, $\tau\tau$, ZZ), and are part of a paper, currently in preparation, presenting the mono-Higgs results produced by the CMS collaboration using data collected in 2016.

11.1 Analysis Prospects

The analysis presented in this document represents the first search for dark matter in the mono-Higgs(WW) channel. Due to the low number of signal events expected, the results are dominated by the statistical uncertainty. This channel will thus benefit from the larger luminosity collected during the full Run 2 of the LHC data taking period ($\sim$ 150 fb^{-1}). Additionally, with more data available, a finer categorization of the events will be possible: the events selected in the $e\mu$ channel can be categorized depending on the charge, the flavour, and the p_T of the leptons ($e^+\mu^-$, $e^-\mu^+$, μ^+e^-, μ^-e^+, where the first lepton presents the larger p_T) and further enhance the statistical sensitivity of the results, as in the case of the SM Higgs boson search. The inclusion of the same-flavour channels (e^+e^-, $\mu^+\mu^-$), not considered in this first analysis to avoid the treatment of the overwhelming Drell-Yan background, may double the number of expected signal events, providing a significant improvement of the search. Finally, the jet categorization used in the Standard Model WW and $h \rightarrow W^+W^-$ analyses can be tested to check if it works also in this beyond-the-standard-model research.

Appendix A

Resumen

La existencia de materia oscura es en la actualidad aceptada por la mayoría de la comunidad científica, gracias a las evidencias cosmológicas que apoyan su presencia en el universo. Entre estas evidencias se encuentra el hecho de que la velocidad de rotación de las galaxias no resulta compatible con las predicciones de la ley de gravitación universal de Newton [1], la detección de lentes gravitacionales (como la producida por el Cúmulo Bala [2]) o las anisotropías observadas en el Fondo Cósmico de Microondas [3] (CMB, del inglés Cosmic Microwave Background). Estas observaciones sugieren la existencia de una gran cantidad de materia no visible, cuya interacción gravitatoria proporciona cohesión a los objetos celestes del cúmulo y actúa como lente proporcionando múltiples imágenes del fondo cósmico. Hasta la fecha no existe ningún indicio sobre cuál es la naturaleza de esta materia oscura o si es sensible a otro tipo de interacciones. Ninguna de las partículas que forman parte del Modelo Estándar (ME), la teoría que hasta la fecha ha descrito con precisión las partículas fundamentales y sus interacciones, es un candidato viable de materia oscura. Distintas búsquedas de materia oscura han sido diseñadas en las últimas décadas, bajo la asunción de estar constituida por un único tipo de partícula que interacciona de alguna manera con la materia conocida. La mayoría de teorías propuestas, como extensiones al ME, consideran como posibles candidatos para ser materia oscura a partículas estables, neutras y que interaccionarían solo a través de la fuerza débil y la fuerza gravitatoria. Estas partículas se conocen como WIMPs, debido a sus siglas en inglés: Weakly-Interacting Massive Particles (Partículas Masivas que Interactúan Débilmente).

Varios experimentos buscan actualmente las WIMPs, bien de manera directa, a través de su potencial interacción con el material de detector, o indirecta, buscando un posible exceso en el flujo de rayos cósmicos que pueda ser explicado por la desintegración de partículas de materia oscura en partículas del ME. Además, en los experimentos situados en colisionadores existen programas específicos de búsqueda de candidatos para ser materia oscura. Al no haberse detectado aún ningún indicio de la misma, en los últimos años los diferentes experimentos han establecido límites de exclusión en los valores de las masa y de las secciones eficaces de producción de los posibles candidatos. El presente trabajo está enfocado en la búsqueda de partículas de materia oscura en colisionadores. Debido a la premisa de que las WIMPs son partículas estables y que interactúan débilmente, una vez que se han producido en la colisión se espera que escapen del detector sin dejar señal. Por ello, una de las técnicas para identificar su producción consiste en seleccionar sucesos que presenten un valor elevado de momento faltante en el plano transverso (p_T^{miss}) al haz debido al desbalance de energía en dicho plano producido por las partículas no detectadas. Además, se requiere la presencia de una partícula del ME en el estado final. Este tipo de

búsquedas se conocen con el nombre de Mono-X, donde la X denota la partícula del ME presente en el estado final. El descubrimiento del bosón de Higgs en el año 2012 [4, 5] fue uno de los mayores hitos científicos de los últimos años, y permitió abrir nuevas vías para la búsqueda de materia oscura, a través del estudio de nuevas signaturas en las que la WIMP se acopla a dicho bosón [6]. La manera más directa para explorar estos posibles acoplamientos es a través de la desintegración del bosón de Higgs en partículas invisibles, siempre y cuando dichas desintegraciones sean permitidas por la cinemática. Esto limita la efectividad de este canal de desintegración al espacio de masas $m_{DM} \leq m_h/2 \approx 60$ GeV. Sin embargo, aquellas signaturas en las que la materia oscura se produce asociada a un bosón de Higgs (llamadas mono-Higgs), dan acceso a un rango de masas de las partículas de materia oscura mucho más amplio, ya que en este caso la masa del bosón de Higgs no es un factor limitante para la producción de materia oscura. Además, mientras que en los modelos Mono-X las partículas del ME son emitidas, principalmente, por radiación de estado inicial (ISR, del inglés Initial State Radiation), en el caso de modelos de Mono-Higgs esto es mucho menos probable. Esto es debido a que el acoplamiento del bosón de Higgs con los quarks ligeros está casi suprimido en el ME, y al hecho de que, al no poseer masa los gluones, su interacción con estos tiene lugar a través de *lazos*. De esta forma, el mediador de la interacción entre las partículas de materia oscura y las de ME debe acoplarse directamente con el bosón de Higgs, permitiendo así obtener información directa sobre el vértice entre el Higgs y la materia oscura [6]. En este trabajo se presentan los resultados de la búsqueda de materia oscura producida junto a un bosón de Higgs en el LHC en colisiones entre protones con una energía de centro de masa de 13 TeV. Los datos usados fueron tomados en el experimento CMS en el año 2016, y corresponden a una luminosidad integrada de 35.9 fb^{-1}. Se han tenido en cuenta dos modelos simplificados siguiendo las recomendaciones del grupo ATLAS/CMS Dark Matter Forum [7] para las búsquedas del Run 2. En ambos casos se ha considerado la desintegración del bosón de Higgs a dos bosones W, cada uno desitegrándose a un leptón (electrón o muón) y un neutrino.

A.1 El Modelo Estándar

El Modelo Estándar de la física de partículas describe de manera efectiva las partículas elementales y sus interacciones. Se trata de una teoría cuántica relativista de campos basada en el grupo de simetrías $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ que describe tres de las cuatro interacciones conocidas: la electromagnética, la débil y la fuerte. La fuerza gravitacional, mucho menos intensa que las otras tres, no se puede acomodar dentro de este modelo. Las piezas fundamentales del ME son las partículas elementales y sus interacciones. Las partículas básicas, a su vez, se clasifican en constituyentes de la materia y propagadores de las interacciones. Los constituyentes de la materia son partículas de spin $\frac{1}{2}$ denominadas fermiones y se dividen en leptones y quarks. Mientras que los leptones interactúan solo a través de la fuerza electro-débil, los quarks sienten también la fuerza fuerte, y poseen un número cuántico adicional conocido como *color*. Hasta el momento los quarks nunca han sido observados en estado libre, sino en estados ligados conocidos como mesones ($q\bar{q}$) o bariones (qqq o $\bar{q}\bar{q}\bar{q}$). Los fermiones se clasifican a su vez en tres familias o generaciones, que se diferencian entre sí por sus masas. Los fermiones de la primera generación son los más ligeros y forman la materia ordinaria, mientras que los fermiones de la segunda y de la tercera generación tienen una masa mayor y pueden ser producidos a energías superiores. Dentro de cada generación se dividen también en *up* y *down*. Los quarks de tipo up tienen carga electromagnetica de $+\frac{2}{3}$, los quarks de tipo down de $-\frac{1}{3}$. Los leptones se separan en leptones cargados, con carga -1, y neutrinos, neutros. Por cada fermión existe un anti-

fermión con la misma masa y el mismo spin, pero carga electromagnética opuesta. En el caso de los neutrinos aún no se sabe si su antipartícula es él mismo o no. Los mediadores de las interacciones fundamentales son partículas de spin 1 llamadas bosones de gauge. El fotón es el mediador de la fuerza electromagnética y se acopla a todas las partículas cargadas, es neutro y no tiene masa. Los gluones son los mediadores de la fuerza fuerte, e interactúan con los quarks y entre ellos mismos. Al igual que los fotones, son partículas neutras y sin masa, pero llevan carga de color. El color tiene tres grados de libertad (rojo, azul y verde), por lo que existen ocho tipos de gluones diferentes. Los bosones W^+, W^- y Z son los mediadores de la fuerza débil e interactúan con los quarks, los leptones y entre ellos mismos. Los bosones W tienen carga +1 o -1, y son uno la antipartícula del otro, mientras que el bosón Z no lleva carga eléctrica y es su propia antipartícula. Son partículas masivas, y adquieren masa a través del mecanismo de ruptura espontánea de la simetría. Este mecanismo introduce un término de masa para los bosones débiles y los fermiones, dejando a los fotones y los gluones sin masa y fue introducido por Higgs [76], Englert y Brout [77]. La consecuencia principal de este mecanismo es la introducción de una nueva partícula, el bosón de Higgs, que fue observada por primera vez en 2012 por los experimentos ATLAS [4] y CMS [5]. No obstante, aunque el ME ha demostrado una gran capacidad predictiva, no se puede considerar como la teoría definitiva en física de partículas, sino un modelo muy eficiente reproduciendo los datos. Queda por incluir la gravedad en su descripción de las interacciones fundamentales, y no es capaz de explicar algunos fenómenos, entre los cuales se encuentra la naturaleza de la materia oscura. En este trabajo se utilizan dos modelos simplificados que amplían el ME como referencia para la búsqueda de materia oscura en el LHC.

A.2 El LHC y el Experimento CMS

Los datos utilizados en este trabajo han sido producidos a través de colisiones entre protones a una energía de centro de masa de 13 TeV por el gran colisionador de hadrones [91, 92] LHC (del inglés Large Hadron Collider). El LHC es un acelerador y colisionador circular de protones o iones pesados que se encuentra entre Suiza y Francia. La alta energía de las colisiones (que puede llegar a un valor nominal de 14 TeV), y la elevada luminosidad instantánea hacen que el LHC sea una máquina ideal para buscar procesos con baja tasa de producción y que involucran partículas de alta masa. En el anillo principal del acelerador, que mide unos 27 km, están situados los cuatro experimentos principales, que detectan las partículas producidas por las colisiones, y almacenan la información necesaria para ser analizada posteriormente: ATLAS, CMS, LHCb y ALICE. El experimento CMS [93, 94] ha sido utilizado para registrar los datos analizados en este trabajo. Estos datos se tomaron durante 2016 y corresponden a una luminosidad integrada total de 35.9 fb^{-1}. CMS es un detector de propósito general, de forma cilíndrica y diseñado para ser compacto y relativamente hermético. Mide 22 m de largo y 14 m de diámetro, y pesa aproximadamente 12500 toneladas. El elemento central del diseño del detector es un solenoide superconductor capaz de producir un campo magnético de 3.8 T en cuyo interior se encuentran diferentes sistemas de detección de partículas. En la región más cercana al punto de colisión se sitúa el sistema de detección de trazas de partículas cargadas, compuesto por un detector de píxeles de silicio de tres capas cilíndricas, rodeado por un detector de tiras de silicio de 10 capas. Cada uno de los dos sistemas de detección se completa por dos tapas circulares, que incrementan su hermeticidad. Los detectores de trazas están rodeados por un calorímetro electromagnético (ECAL) de cristales de tungstato de plomo (PbWO4) y un calorímetro hadrónico (HCAL) basado en centelladores, que miden la energía de las

partículas electromagnéticas y hadrónicas, respectivamente. En la parte más externa del detector, se encuentra el hierro de retorno del campo magnético, equipado con un sistema de detectores gaseosos de muones basados en diferentes tecnologías, dependiendo de la región del detector. La zona central del sistema de detección de muones está formada por cámaras de deriva (DT, del inglés Drift Tube), mientras que las tapas del cilindro están formadas por cámaras de tiras catódicas (CSC, del inglés Cathode Strip Chamber). En ambos casos DT y CSC son complementadas por cámaras de tiras resistivas (RPC, del inglés Resistive Plate Chambers). Debido a la elevada frecuencia de interacción de colisiones entre protones (40 MHz, correspondientes a un choque cada 25 ns), la velocidad y el espacio de almacenamiento no son suficientes para guardar todos los sucesos producidos. Es por lo que un sistema de selección de sucesos, llamado *trigger*, se encarga de elegir y guardar hasta 2000 sucesos por segundo, los más interesantes desde el punto de vista de la física.

A.3 Reconstrucción de Objetos

Los diferentes detectores que componen CMS producen señales eléctricas cuando las partículas los atraviesan, que se interpretan como interacciones en los pixeles o piezas de silicio del trazador o en las cámaras de muones, o como depósitos de energía en los calorímetros. Para interpretar dichas señales eléctricas en términos de partículas y sus propiedades se utiliza un sistema de algoritmos que explota la información procedente de todos los detectores que forman parte de CMS. En el caso de la búsqueda de mono-Higgs a dos bosones W, se seleccionan dos leptones de distinto sabor, un electrón y un muón, en el estado final. Esto requiere tener un eficiente sistema de detección, reconstrucción e identificación de los leptones. Los leptones que provienen de la desintegración de un bosón W tienden a estar aislados de otras partículas producidas en el mismo suceso, mientras que hadrones que se hayan identificado erróneamente como leptones, o leptones procedentes de la desintegración de un quark pesado se encuentran típicamente dentro de un jet. Para discriminar de manera eficaz entre estos dos tipos de leptones se introduce el concepto de *aislamiento*. El aislamiento se define como la suma escalar de los momentos de las partículas producidas en un cono de un cierto radio alrededor del leptón. Si esa suma es más grande que una cierta fracción del momento del leptón, este se considera como no aislado, de manera que solo leptones que presenten un bajo valor de aislamiento son seleccionados para el análisis. Debido a la presencia de dos neutrinos procedentes de las desintegraciones leptónicas de los bosones W, y de dos partículas de materia oscura, los sucesos seleccionados presentan una gran cantidad de momento faltante en el plano transverso (p_T^{miss}). Este observable se calcula como la magnitud de la suma vectorial de los momentos de todas las partículas del suceso, cambiada de signo. Finalmente, aunque no haya selección en el número de jets reconstruidos en un suceso, los jets reconstruidos como procedentes de la desintegración de un quark b (b-jets) son relevantes para este análisis. Rechazar sucesos que contengan b-jets permite reducir fuertemente la contaminación del fondo de top. Para la identificación de b-jets se utilizan técnicas de análisis multivariable que consideran un gran número de variables cinemáticas de los b-jets (como el tiempo de vuelo y la vida media del quark b, más larga con respecto a la de los quarks ligeros) para distinguirlos de otros jets.

A.4 Análisis de Datos

Los modelos usados como referencia para la búsqueda de materia oscura siguen las recomendaciones del ATLAS-CMS Dark Matter Forum [7]. Dos extensiones del ME han

sido estudiadas. El modelo Z'-2HDM [124] introduce un nuevo mediador vectorial Z' que se produce a través de la aniquilación de dos quarks y que se desintegra a un bosón de Higgs y a una partícula pseudoscalar A. La partícula A se desintegra finalmente en un par de partículas de materia oscura. En el segundo modelo, Z' bariónico, un mediador Z' leptofóbico puede irradiar un bosón de Higgs y luego desintegrarse en un par de partículas de materia oscura. En ambos casos se considera la desintegración del bosón de Higgs a dos bosones W, cada uno desintegrandose en un leptón y un neutrino. El canal de desintegración del bosón de Higgs a dos W posee una fracción de desintegración del 21.5%, después del canal a quarks b. El canal de desintegración leptónico de los bosones W reduce en parte el número de eventos disponibles, pero por otro lado asegura tener un estado final limpio, con una contaminación de fondos relativamente pequeña y un buen control de las incertidumbres sistemáticas. Entre los procesos del ME que producen un estado final similar al de la señal se encuentran la producción de un bosón de Higgs desintegrándose a dos bosones W y la producción de dos bosones W tal y como predice el ME. La desintegración de pares de quarks top a dos leptones, o la producción de un quark top y un bosón W difieren de la señal por la presencia de b-jets en el estado final, pero debido a su gran sección eficaz y que la identificación de b-jets no es perfecta, estos procesos representan una fuente importante de fondo. El proceso Drell-Yan, cuando el Z o el γ^* se desintegran a pares de leptones τ, y ambos τ se desintegran leptónicamente ($\tau \to \ell\bar{\nu}_\ell\nu_\tau$), produce un electrón, un muón y cuatro neutrinos, los cuales generan una gran cantidad de p_T^{miss}. Otro fondo importante son aquellos sucesos con un leptón real y un leptón no reconstruido o no indentificado correctamente (denominados leptones *non-prompt*). Finalmente, otros procesos como dibosones o tribosones, representan fondos si producen al menos dos leptones en el estado final y los leptones adicionales no son reconstruidos o caen en una región fuera de la aceptancia del detector.

A.4.1 Selección de Sucesos

Para definir un espacio de fases enriquecido en sucesos de señal el primer requisito es pedir sucesos seleccionados por un *trigger* de uno o dos leptones (en ese caso un electrón y un muón). Dichos leptones tienen que ser de carga opuesta y tener un momento transverso (p_T) mayor que 25 GeV (el más energético) y 20 GeV (el menos energético). Para reducir la contaminación de sucesos con leptones *non-prompt*, el momento transverso del sistema de los dos leptones tiene que ser mayor de 30 GeV, mientras que sucesos con un tercer leptón son rechazados para suprimir sucesos de tribosones. La masa invariante del sistema de los dos leptones tiene que ser mayor que 12 GeV para rechazar resonancias ligeras. Para reducir la contaminación de sucesos Drell-Yan se pide un p_T^{miss} mayor que 20 GeV y una masa transversa del sistema de los dos leptones y p_T^{miss} ($m_T^{\ell\ell E_T^{miss}}$) mayor que 40 GeV. Finalmente, procesos que involucren quarks top son rechazados vetando sucesos que contengan b-jets. Además de estas selecciones, comunes a los análisis de WW y $h \to W^+W^-$, para el caso del canal mono-Higgs, se ha realizado una selección más estricta tras estudiar cuidadosamente las distribuciones cinemáticas de la señal y de los fondos principales. La masa invariante de los dos leptones procedentes de la desintegración de un bosón de Higgs alcanza un máximo en valores relativamente bajos, de manera que se seleccionan sucesos con $m_{\ell\ell} < 76$ GeV. Además, debido al alto momento transverso del bosón de Higgs que retrocede frente a las partículas de materia oscura, los dos leptones suelen estar más cerca entre ellos, con respecto a los producidos por procesos del ME, de manera que se guardan los sucesos con los dos leptones que tengan una distancia en el plano (η,ϕ) ($\Delta R(\ell,\ell)$) menor de 2.5.

A.4.2 Estudio de Fondos

La contaminación de los fondos en la región de señal se estima directamente de los datos siempre y cuando esto es posible, como en el caso de procesos que producen leptones *non-prompt*: la probabilidad de que un leptón non-prompt sea identificado como un leptón prompt (fake rate) es medida en un espacio de fases enriquecido en jets, en el cual no se espera una presencia significativa de leptones prompt. El fake rate se aplica a una muestra de datos en la cual un leptón pasa las selecciones del análisis, y el otro no. En el caso del proceso top, la estimación de la contaminación en la región de señal se basa en simulación, corregida por cualquier discrepancia con los datos. Esta corrección se realiza en dos pasos: en el primero se aplican factores de corrección a la eficiencia de identificación de b-jets (que puede ser diferente entre datos y simulación), mientras que en el segundo se define una región enriquecida en sucesos de top pidiendo la presencia de al menos un b-jet, que se usa para fijar la normalización del proceso. De la misma manera los procesos WW y Drell-Yan son estimados en simulación, y sus normalizaciones son ajustadas a los datos en dos regiones de control distintas. En el caso del proceso WW dicha región se define relajando el corte en $\Delta R(\ell, \ell)$ y pidiendo que la masa invariante de los dos leptones sea mayor que 76 GeV, mientras que en el caso del proceso Drell-Yan se pide $m_T^{\ell\ell E_T^{miss}} < 40$ GeV. Los procesos WZ y Wγ^* producen tres leptones, y pueden contaminar la región de señal cuando uno de ellos no se reconstruye o se produce en una zona del detector fuera de la aceptancia. Para estimar el nivel de contaminación, una única muestra de simulación es utilizada para los dos procesos, aunque para fijar las normalizaciones sean definidas dos regiones de control distintas. El proceso Wγ^* se normaliza a los datos en una región con tres leptones, donde se requiere que dos leptones del mismo sabor y carga opuesta (procedentes de la desintegración del γ^*) tengan masa invariante menor que 4 GeV. Para el proceso WZ, los dos leptones producidos por la desintegración del bosón Z tienen que tener una masa invariante entre 80 GeV y 110 GeV. Finalmente, los demás fondos, principalmente la producción de un bosón de Higgs, o producción de sucesos de dibosones o tribosones, son estimados directamente utilizando simulaciones, ya que son fondos menores o es imposible definir una región de control con una pureza y una cantidad de datos suficiente.

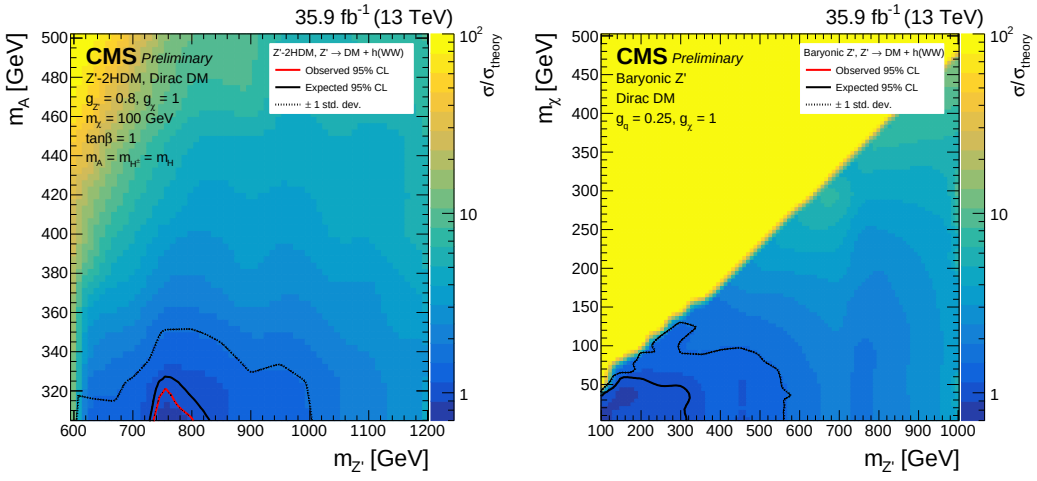
A.4.3 Extracción de la Señal e Incertidumbres Sistemáticas

La región de señal presenta una importante contaminación de otros procesos que consideramos como fondos. Existen diferentes variables cinemáticas que muestran que es posible mejorar la discriminación entre señal y fondo, aunque ninguna pueda marcar la diferencia individualmente. Por lo tanto, la extracción de la señal se ha realizado utilizando técnicas de análisis multivariable, en particular utilizando la técnica de boosted decision trees (BDT). Se ha entrenado un BDT específico para cada uno de los dos modelos usando como variables de entrada la masa transversa, los momentos de los leptones, la energía faltante, distintas variables angulares entre leptones y p_T^{miss}, y la masa invariante de los dos leptones. Para maximizar el poder de discriminación entre señal y fondo en cada BDT y reducir el sobre-entrenamiento, se ha realizado un estudio sistemático y detallado de los parámetros de entrenamiento, como el número de *trees*, la profundidad de los *trees*, la velocidad de aprendizaje y el tamaño mínimo de los nudos. Además, ya que para cada modelo se utiliza un solo BDT para todos los puntos de masa considerados, para su entrenamiento se han utilizados varias muestras de señal. Esto asegura que los BDT sepan distinguir todos los puntos de masa de señal de los fondos principales, sin dar una respuesta preferente a una masa en particular. Para el análisis final se ha realizado un ajuste por bines para maximizar la verosimilitud de las distribuciones de salida del BDT y evaluar la cantidad

de sucesos de señal y de fondo resultantes. El tamaño de cada bin de la distribución del BDT se ha elegido de manera que la significancia $\frac{S}{\sqrt{S+B}}$ sea máxima en cada bin. Para la realización del ajuste se han considerado las diferentes fuentes de incertidumbre que pueden afectar a la medida. El bajo número de sucesos de señal esperado y el bajo número de sucesos de datos en los bines más significativos de la distribución hacen que la incertidumbre estadística domine los resultados. Otras fuentes de incertidumbre adicionales, tanto experimentales como teóricas, han sido consideradas. Entre ellas, la incertidumbre de la luminosidad integrada, de la eficiencia de trigger, de la eficiencia de reconstrucción e identificación de leptones, de la escala de energía y momento de los leptones y jets, del modelado del p_T^{miss} y de la eficiencia de identificación de b-jets, que entran en la categoría de incertidumbres experimentales. En el caso de las incertidumbres teóricas, se tienen en cuenta las limitaciones en el conocimiento de las funciones de distribución de los partones (PDF, del inglés parton distribution functions), las correcciones de orden superior faltantes en el cálculo de las secciones eficaces y en el modelado de las variables cinemáticas, y el modelado de los sucesos secundarios en el choque de protones y de la emisión de partones.

A.5 Resultados

Los resultados de los ajustes no muestran diferencias significativas entre la cantidad de sucesos esperados por el ME y los datos, de manera que son interpretados como límites superiores en la sección eficaz de producción de materia oscura en el LHC. En particular, fijando la sección eficaz de producción de diferentes puntos de masa para cada uno de los dos modelos considerados, ha sido posible excluir para el modelo Z'-2HDM puntos de masa del mediador Z' entre 740 GeV y 800 GeV y del pseudoescalar A menor que 320 GeV, mientras que no ha sido posible excluir ningún punto de masa para el modelo Z' bariónico, como se muestra en la Figura A.1.



(a) $m_{Z'}$ vs m_A limits.

(b) $m_{Z'}$ vs m_χ limits.

Figure A.1: Límites de exclusión para el modelo Z'-2HDM (izquierda) y para el modelo Z' bariónico (derecha). Los límites se muestran en función de $m_{Z'}$ y m_A para el modelo Z'-2HDM y en función de $m_{Z'}$ y m_χ para el modelo Z' bariónico. La línea negra representa los límites esperados, mientras que la línea roja representa los límites observados.

Appendix B

BDT Parameters Studies

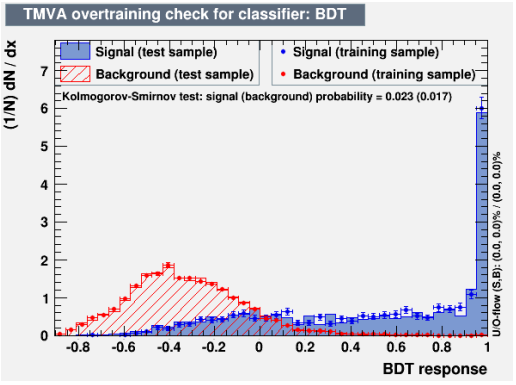
To select the parameters of the MVA training, a scan of each of them has been performed, separately for the adaptive boost and the gradient boost training method. In the study, the sample simulating the Z'-2HDM model, with $m_{Z'} = 600$ GeV and $m_A = 300$ GeV is considered as signal. All the backgrounds considered in the analysis are used here, too. Namely, they are ZH, Higgs, WW and Top. The selections used correspond to that used for the eμ channel. For each training algorithm, each parameter is moved keeping the others fixed to the nominal values documented in Section 9.1. They are recalled in Table B.1.

Gradient Boost	Adaptive Boost
NTrees = 500	NTrees = 500
MinNodeSize = 1.5%	MinNodeSize = 0.5%
BoostType = Grad	BoostType = AdaBoost
Shrinkage = 0.05	AdaBoostBeta = 0.1
GradBaggingFraction = 0.5	
nCuts = 500	nCuts = 500
MaxDepth = 2	MaxDepth = 2

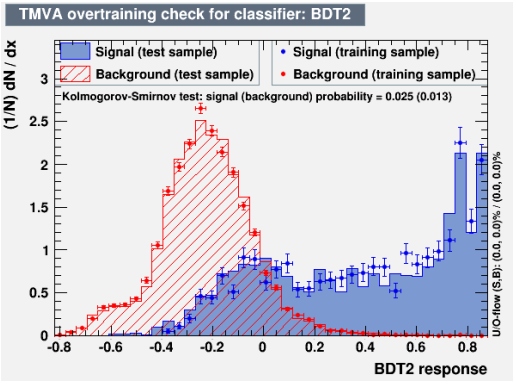
Table B.1: Summary of BDT training options used in the analysis.

B.1 Number of Trees

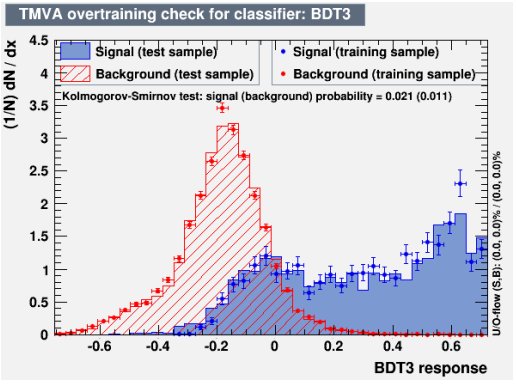
B.1.1 Adaptive Boost



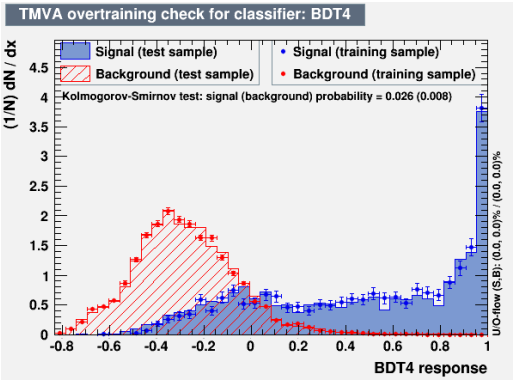
(a) nTrees = 250.



(b) nTrees = 1000.



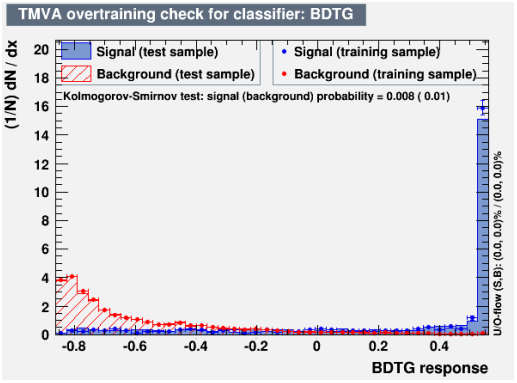
(c) nTrees = 2000.



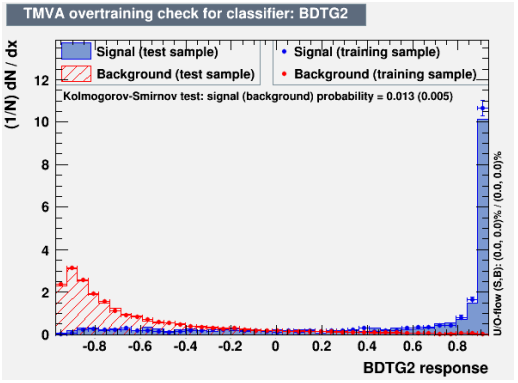
(d) nTrees = 500.

Figure B.1: Effect of the number of trained trees on the output of the BDT forest using the adaptive boost algorithm.

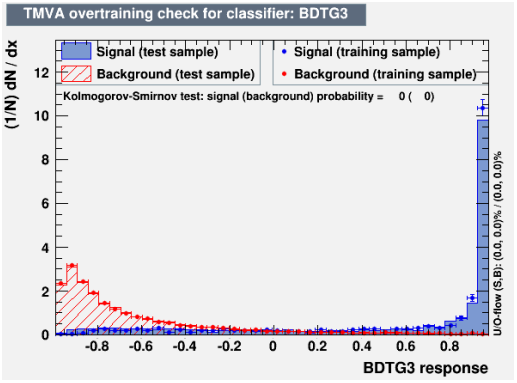
B.1.2 Gradient Boost



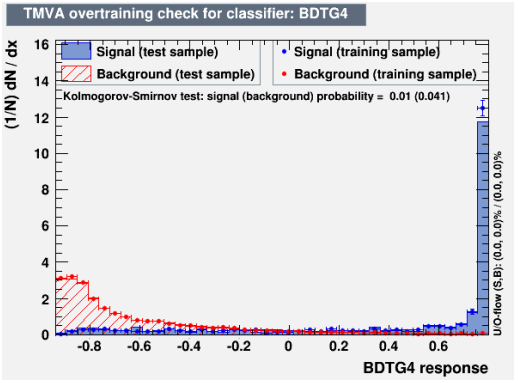
(a) nTrees = 250.



(b) nTrees = 1000.



(c) nTrees = 2000.

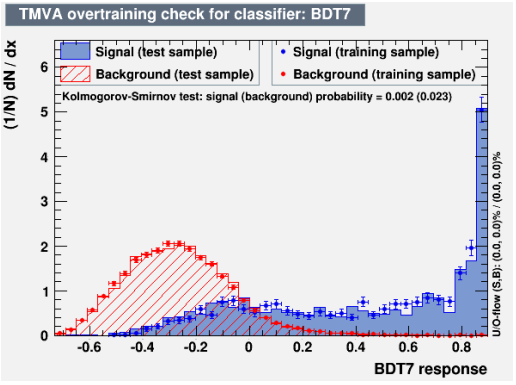


(d) nTrees = 500.

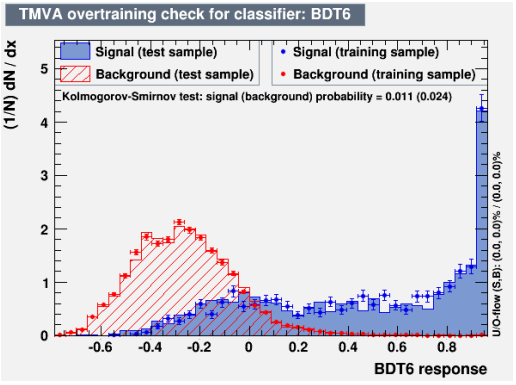
Figure B.2: Effect of the number of trained trees on the output of the BDT forest using the gradient boost algorithm.

B.2 Minimum Node Size

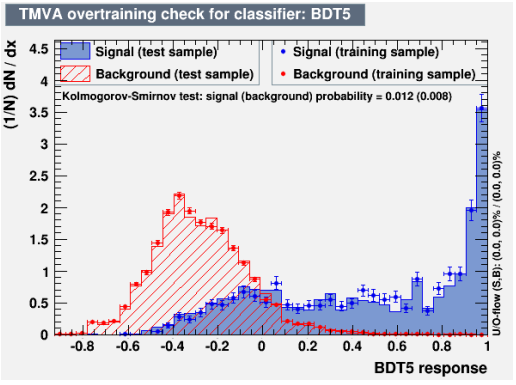
B.2.1 Adaptive Boost



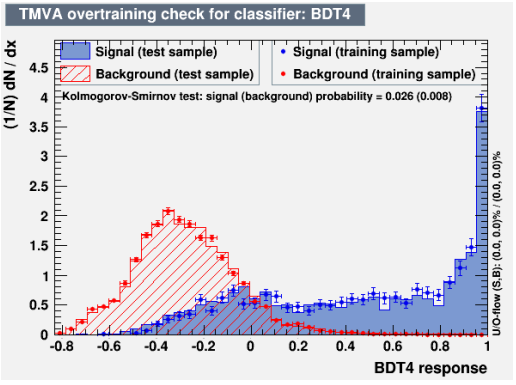
(a) nodeSize = 5%.



(b) nodeSize = 2.5%.



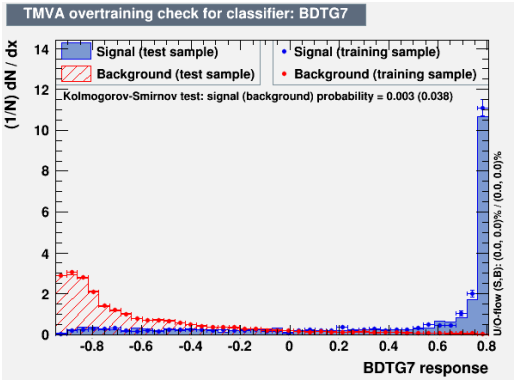
(c) nodeSize = 0.25%.



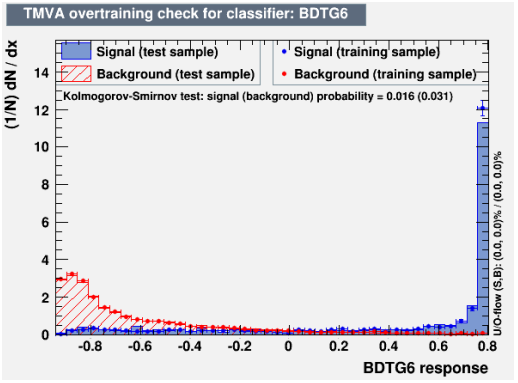
(d) nodeSize = 0.5%.

Figure B.3: Effect of the minimum node size on the output of the BDT forest using the adaptive boost algorithm.

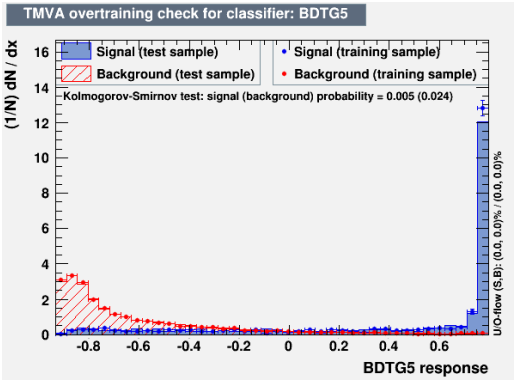
B.2.2 Gradient Boost



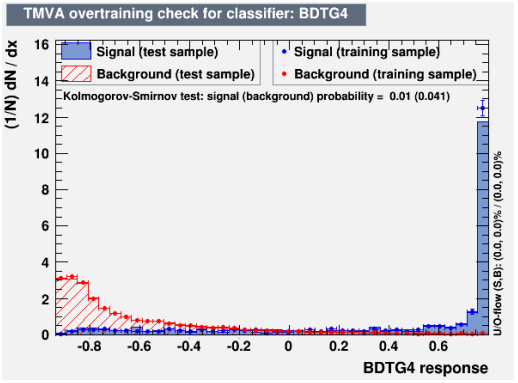
(a) nodeSize = 5%.



(b) nodeSize = 2.5%.



(c) nodeSize = 0.5%.

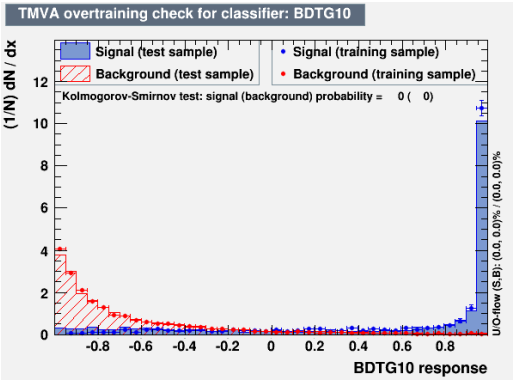


(d) nodeSize = 1.5%.

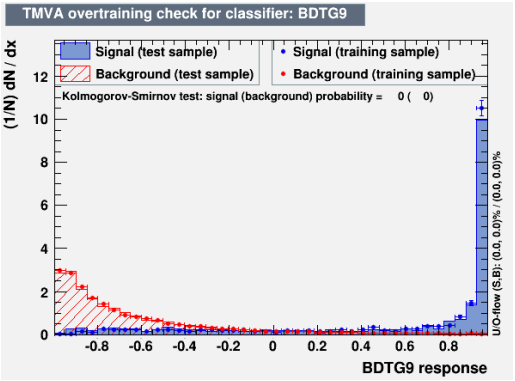
Figure B.4: Effect of the minimum node size on the output of the BDT forest using the gradient boost algorithm.

B.3 Shrinkage

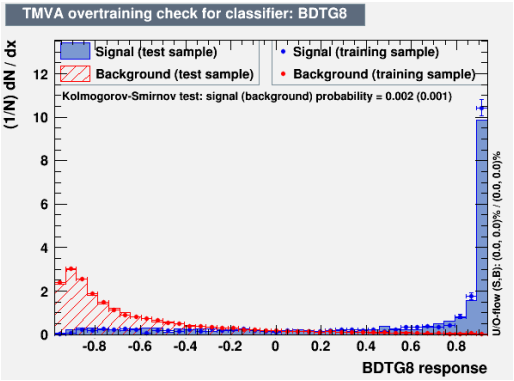
B.3.1 Gradient Boost



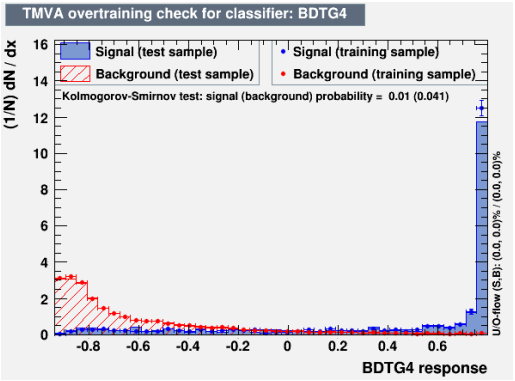
(a) Shrinkage = 0.5.



(b) Shrinkage = 0.2.



(c) Shrinkage = 0.1.

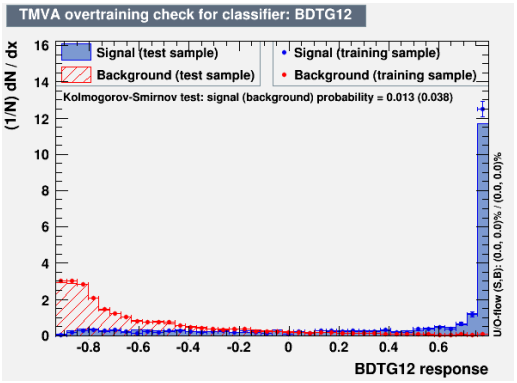


(d) Shrinkage = 0.05.

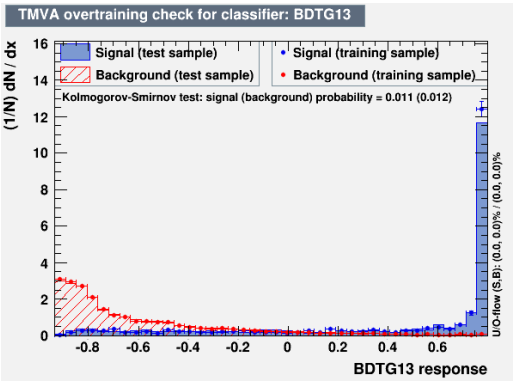
Figure B.5: Effect of the shrinkage on the output of the BDT forest using the gradient boost algorithm.

B.4 Bagging Fraction

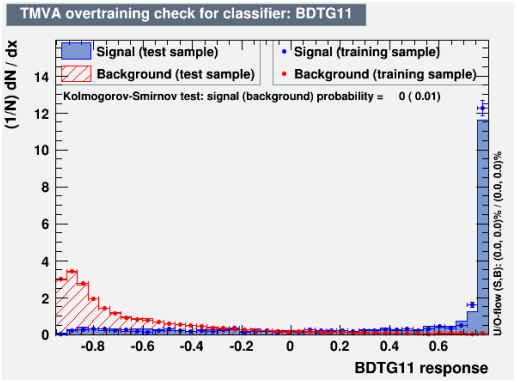
B.4.1 Gradient Boost



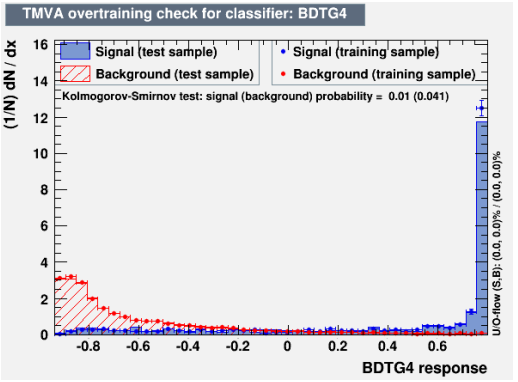
(a) baggingFraction = 0.75.



(b) baggingFraction = 1.



(c) baggingFraction = 0.25.

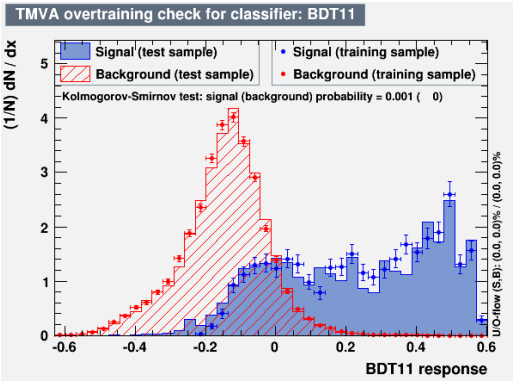


(d) baggingFraction = 0.5.

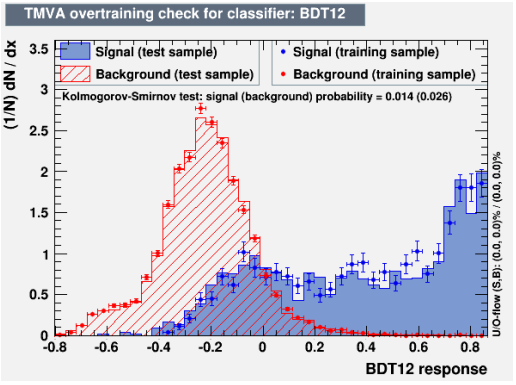
Figure B.6: Effect of the bagging fraction on the output of the BDT forest using the gradient boost algorithm.

B.5 Adaptive Boost Beta

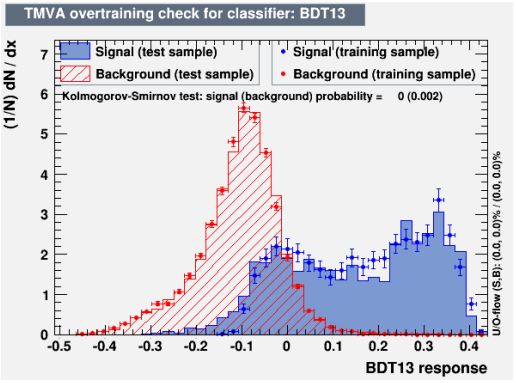
B.5.1 Adaptive Boost



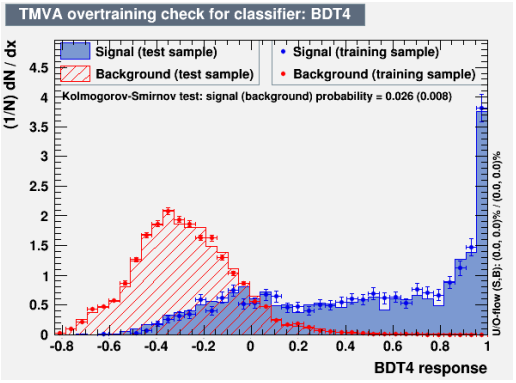
(a) adaBoostBeta = 0.5.



(b) adaBoostBeta = 0.2.



(c) adaBoostBeta = 0.8.

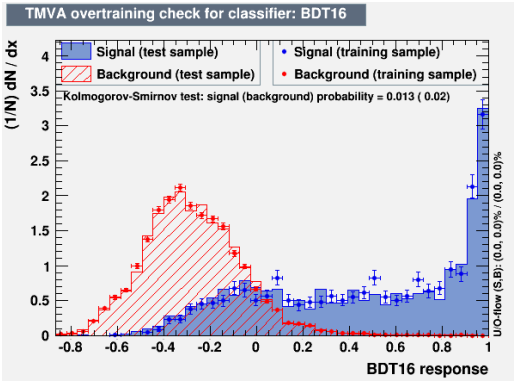


(d) adaBoostBeta = 0.1.

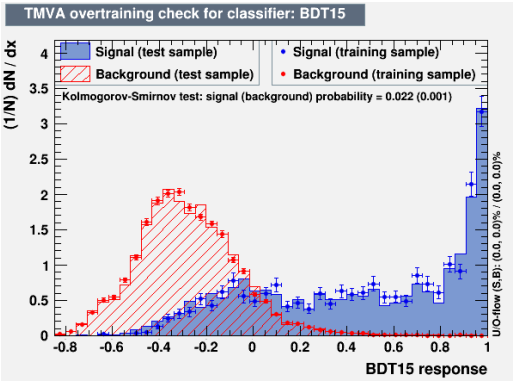
Figure B.7: Effect of the adaptive boost beta on the output of the BDT forest using the adaptive boost algorithm.

B.6 Number of Cuts

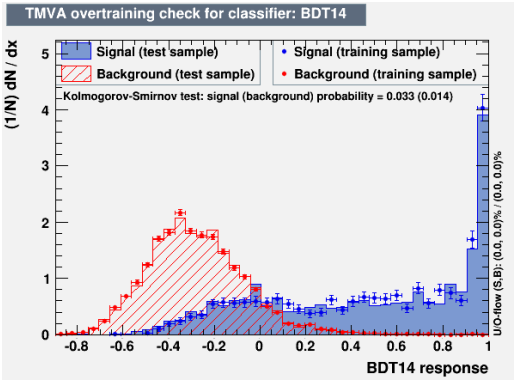
B.6.1 Adaptive Boost



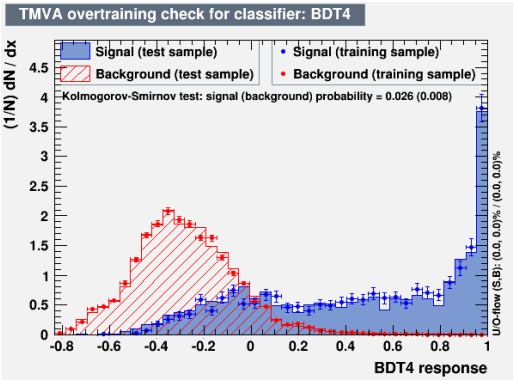
(a) nCuts = 10000.



(b) nCuts = 1000.



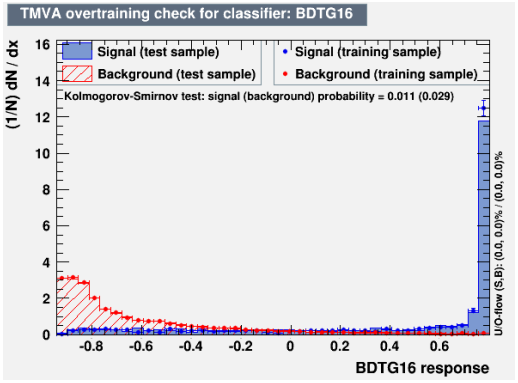
(c) nCuts = 100.



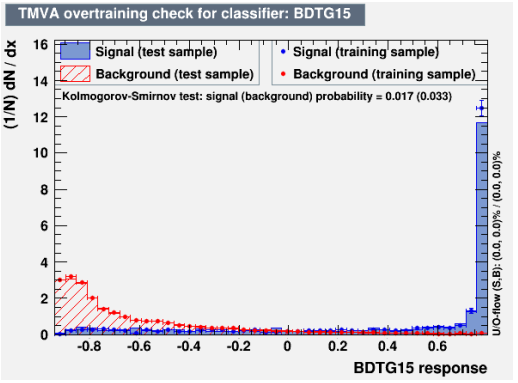
(d) nCuts = 500.

Figure B.8: Effect of the number of cuts on the output of the BDT forest using the adaptive boost algorithm.

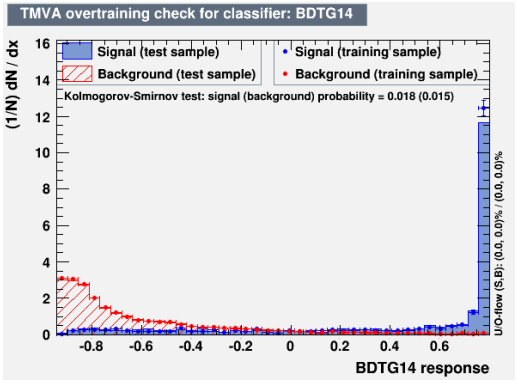
B.6.2 Gradient Boost



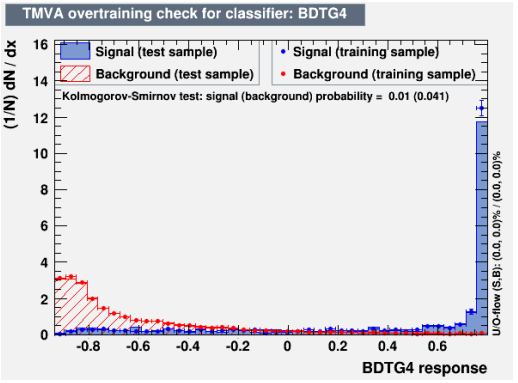
(a) nCuts = 10000.



(b) nCuts = 1000.



(c) nCuts = 100.

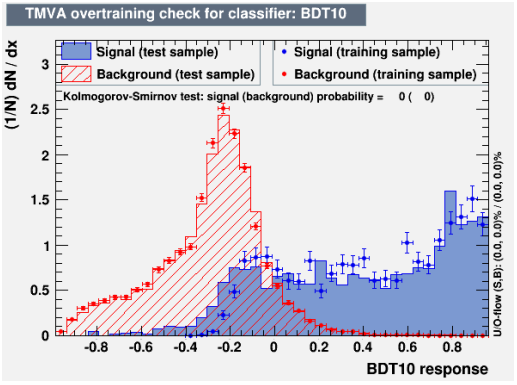


(d) nCuts = 500.

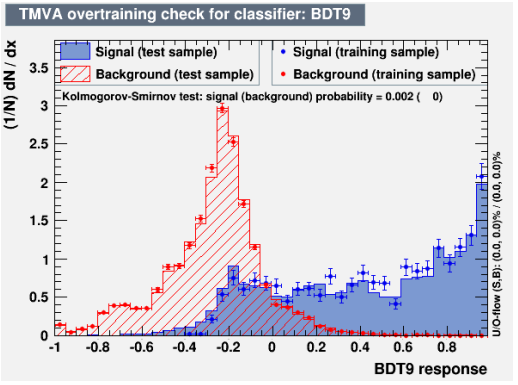
Figure B.9: Effect of the number of cuts on the output of the BDT forest using the gradient boost algorithm.

B.7 Maximum Tree Depth

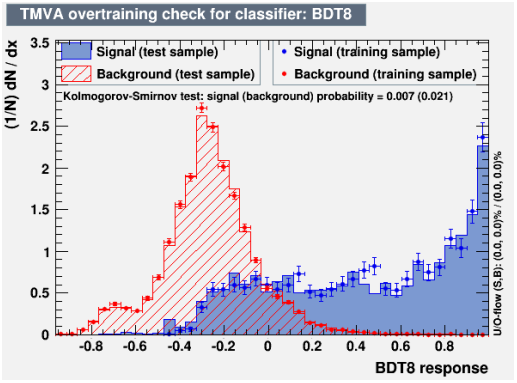
B.7.1 Adaptive Boost



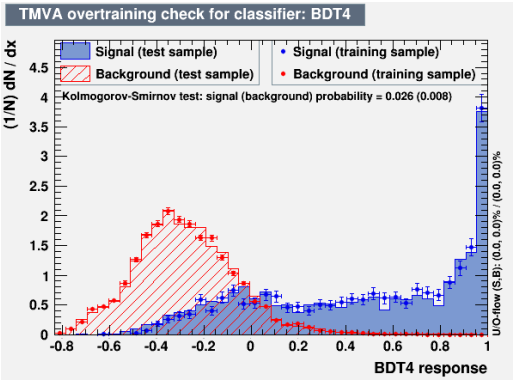
(a) maxDepth = 5.



(b) maxDepth = 4.



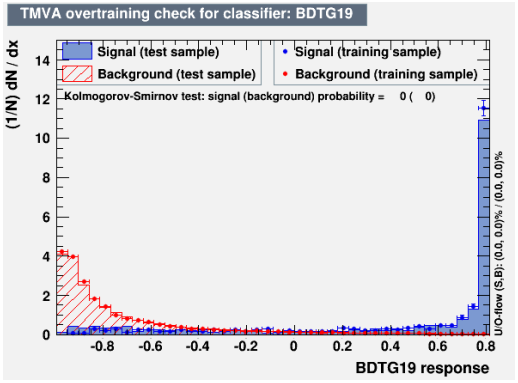
(c) maxDepth = 3.



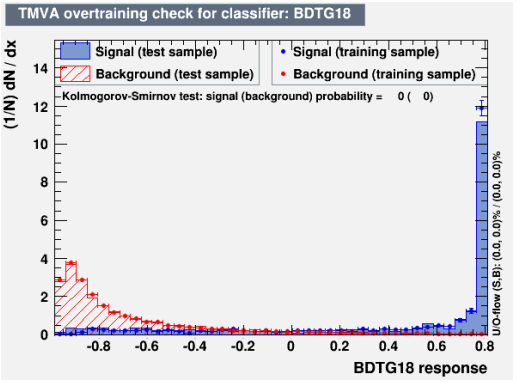
(d) maxDepth = 2.

Figure B.10: Effect of the maximum tree depth on the output of the BDT forest using the adaptive boost algorithm.

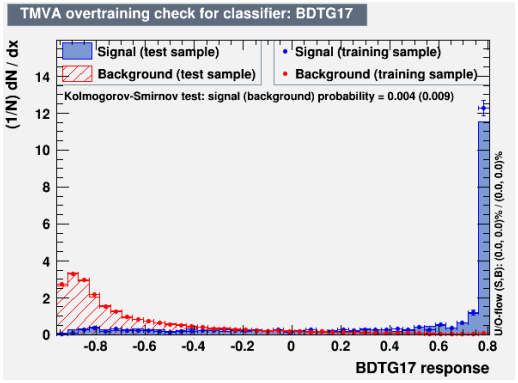
B.7.2 Gradient Boost



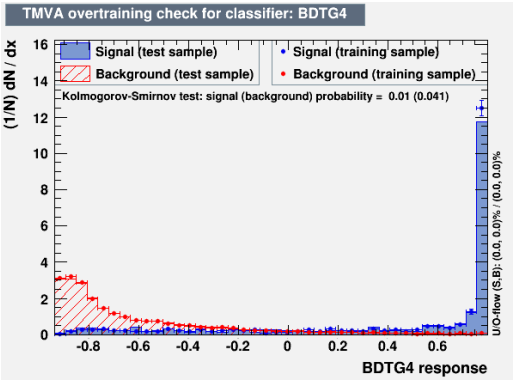
(a) maxDepth = 5.



(b) maxDepth = 4.



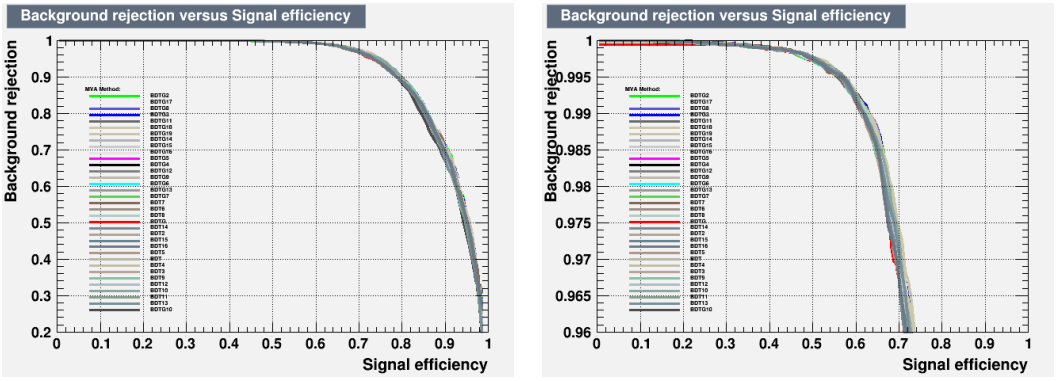
(c) maxDepth = 3.



(d) maxDepth = 2.

Figure B.11: Effect of the maximum tree depth on the output of the BDT forest using the gradient boost algorithm.

B.7.3 ROC Curves



(a) ROC curves.

(b) ROC curves - zooming the high background rejection region.

Figure B.12: Effect of the maximum tree depth on the ROC curve for the two training algorithms.

Appendix C

Pre-Fit Tables

	monoH-1200-300	monoH-ZB-500-1	ggWW	ggH-hww	qqH-hww	Vg	VVV	top	ggZH-hww
	3.01	29.60	543.87	509.62	49.93	133.47	15.18	4448.25	2.44
BR-hWW (lnN)	1.004	1.004	-	1.004	1.004	-	-	-	1.004
BR-htt (lnN)	-	-	-	-	-	-	-	-	-
QCDscale-WH (lnN)	-	-	-	-	-	-	-	-	-
QCDscale-ZH (lnN)	-	-	-	-	-	-	-	-	-
QCDscale-ggH (lnN)	-	-	-	1.0854	-	-	-	-	-
QCDscale-ggH-monoH (lnN)	-	-	-	-	-	-	-	-	-
QCDscale-ggZH (lnN)	-	-	1.027	1.027	-	-	-	-	1.37
QCDscale-gg-accept (lnN)	-	-	-	-	0.997, 1.004	-	-	-	1.027
QCDscale-qqH (lnN)	-	-	-	-	1.007	-	-	-	-
QCDscale-qqbar-accept (lnN)	-	-	-	1.03	1.03	-	-	-	-
UE (lnN)	-	-	-	1.031	-	-	-	-	1.15
pdf-gg (lnN)	-	-	1.005	1.005	-	-	-	-	1.005
pdf-gg-accept (lnN)	-	-	-	-	1.021	-	-	-	-
pdf-qqbar (lnN)	-	-	-	-	1.011	-	-	-	-
pdf-qqbar-accept (lnN)	-	-	-	-	-	-	-	-	-
CMS-fake-ele-stat-hww (shape)	-	-	-	-	-	-	-	-	-
CMS-fake-mu-hww (shape)	-	-	-	-	-	-	-	-	-
CMS-fake-mu-stat-hww (shape)	-	-	-	-	-	-	-	-	-
CMS-scale-e (shape)	1.000/0.999	1.002/0.996	1.001/0.999	1.008/0.992	1.008/0.993	1.031/0.982	0.997/0.999	1.000/0.999	1.003/0.997
CMS-scale-j (shape)	0.996/1.004	0.994/1.005	0.993/1.008	0.991/1.010	0.978/1.023	0.997/1.017	0.988/1.011	0.974/1.027	0.992/1.009
CMS-scale-m (shape)	0.999/1.000	1.001/0.999	1.001/0.999	1.004/0.997	1.006/0.994	1.000/0.959	1.002/1.002	1.001/0.999	1.002/0.997
CMS-scale-met (shape)	1.000/1.000	0.999/1.003	0.998/1.001	0.995/1.003	1.000/1.000	1.007/1.008	0.999/0.994	0.997/1.001	1.000/0.999
CMS-trigger (shape)	1.002/0.998	1.003/0.997	1.003/0.997	1.003/0.997	1.003/0.997	1.004/0.997	1.003/0.997	1.003/0.997	1.003/0.997
CMS-WWqscaleMonoH (shape)	-	-	-	-	-	-	-	-	-
CMS-WWresumMonoH (shape)	-	-	-	-	-	-	-	-	-
CMS-btag-bc (shape)	0.999/1.001	0.999/1.001	0.999/1.001	0.999/1.001	0.997/1.003	0.999/1.001	0.995/1.005	0.941/1.061	-
CMS-btag-udsg (shape)	0.994/1.006	0.996/1.004	0.996/1.004	0.995/1.005	0.992/1.008	0.994/1.006	0.989/1.011	0.994/1.006	-
CMS-hww-WZScale (lnN)	-	-	-	-	-	-	-	-	-
CMS-hww-WgStarScale (lnN)	-	-	-	-	-	-	-	-	-
CMS-idiso-ele (shape)	1.012/0.988	1.008/0.992	1.007/0.993	1.007/0.993	1.007/0.993	1.010/0.990	1.008/0.992	1.007/0.993	1.008/0.992
CMS-idiso-mu (shape)	1.013/0.988	1.011/0.989	1.011/0.989	1.011/0.989	1.011/0.989	1.011/0.989	1.011/0.989	1.011/0.989	1.011/0.989
CMS-lumi (lnN)	1.025	1.025	-	1.025	1.025	1.025	1.025	-	1.025
CMS-toppTtw (shape)	-	-	-	-	-	-	-	1.014/1.000	-
CMS-tttwTh (shape)	-	-	-	-	-	-	-	1.015/0.981	-
PS (lnN)	-	-	-	1.02	1.02	-	-	-	-
fake-syst (lnN)	-	-	-	-	-	-	-	-	-
kfactggww (lnN)	-	-	1.15	-	-	-	-	-	-

Table C.1: Pre-fit nuisances (first part).

	WZgS-L	WH-hww	WZgS-H	DY	H-htt	Fake	bbH-hww	VZ	WW	ZH-hww
	62.68	19.46	228.29	366.57	6.32	664.45	3.21	9.05	3958.14	7.07
BR-hWW (lnN)	-	1.004	-	-	-	-	1.004	-	-	1.004
BR-htt (lnN)	-	-	-	-	1.005	-	-	-	-	-
QCDscale-WH (lnN)	-	0.993, 1.005	-	-	-	-	-	-	-	-
QCDscale-ZH (lnN)	-	-	-	-	-	-	-	-	-	0.969, 1.038
QCDscale-ggH (lnN)	-	-	-	-	0.919, 1.076	-	-	-	-	-
QCDscale-ggH-monoH (lnN)	-	-	-	-	-	-	-	-	-	-
QCDscale-ggZH (lnN)	-	-	-	-	-	-	-	-	-	-
QCDscale-gg-accept (lnN)	-	-	-	-	1.027	-	-	-	-	-
QCDscale-qqH (lnN)	-	-	-	-	-	-	-	-	-	-
QCDscale-qqbar-accept (lnN)	-	1.05	-	-	-	-	-	1.029	-	1.04
UE (lnN)	-	-	-	-	-	-	-	-	-	-
pdf-gg (lnN)	-	-	-	-	1.031	-	-	-	-	-
pdf-gg-accept (lnN)	-	-	-	-	1.005	-	-	-	-	-
pdf-qqbar (lnN)	-	1.019	-	-	-	-	-	1.04	-	1.016
pdf-qqbar-accept (lnN)	-	1.007	-	-	-	-	-	1.005	-	1.012
CMS-fake-ele-stat-hww (shape)	-	-	-	-	-	1.082/0.926	-	-	-	-
CMS-fake-mu-hww (shape)	-	-	-	-	-	0.967/1.016	-	-	-	-
CMS-fake-mu-stat-hww (shape)	-	-	-	-	-	1.022/0.979	-	-	-	-
CMS-scale-e (shape)	1.009/0.990	1.002/0.996	1.004/0.997	1.007/0.990	0.997/0.999	-	1.009/0.979	1.005/0.994	1.000/0.998	1.005/0.995
CMS-scale-j (shape)	0.986/1.013	0.986/1.006	0.992/1.012	0.980/1.031	0.966/0.951	-	0.968/1.019	1.007/1.009	0.994/1.006	0.989/1.011
CMS-scale-m (shape)	1.006/0.995	1.003/0.998	1.003/0.997	1.004/0.999	0.991/0.972	-	0.997/0.995	1.013/0.984	1.001/0.998	1.003/0.996
CMS-scale-met (shape)	0.996/1.004	0.982/1.005	0.998/1.002	1.138/0.930	0.986/0.937	-	0.981/1.013	1.044/0.988	0.994/1.005	0.999/1.002
CMS-trigger (shape)	1.004/0.997	1.003/0.997	1.004/0.997	1.003/0.997	1.003/0.997	-	1.003/0.997	1.003/0.997	1.003/0.997	1.003/0.997
CMS-WWqscaleMonoH (shape)	-	-	-	-	-	-	-	-	0.997/0.999	-
CMS-WWresumMonoH (shape)	-	-	-	-	-	-	-	-	1.011/0.953	-
CMS-btag-bc (shape)	0.998/1.002	0.995/1.005	0.998/1.002	0.997/1.003	0.999/1.001	-	0.989/1.011	0.996/1.004	0.999/1.001	0.995/1.005
CMS-btag-udsg (shape)	0.994/1.006	0.991/1.009	0.993/1.007	0.991/1.009	0.991/1.009	-	0.997/1.003	0.991/1.009	0.996/1.004	0.991/1.009
CMS-hww-WZScale (lnN)	-	-	1.16	-	-	-	-	-	-	-
CMS-hww-WgStarScale (lnN)	1.25	-	-	-	-	-	-	-	-	-
CMS-idiso-ele (shape)	1.008/0.992	1.007/0.993	1.007/0.993	1.008/0.992	1.007/0.993	-	1.007/0.993	1.008/0.992	1.007/0.993	1.008/0.992
CMS-idiso-mu (shape)	1.011/0.989	1.011/0.989	1.011/0.989	1.011/0.989	1.012/0.988	-	1.011/0.989	1.011/0.989	1.011/0.989	1.011/0.989
CMS-lumi (lnN)	1.025	1.025	1.025	-	1.025	-	1.025	1.025	-	1.025
CMS-toppTrw (shape)	-	-	-	-	-	-	-	-	-	-
CMS-tttwTh (shape)	-	-	-	-	-	-	-	-	-	-
PS (lnN)	-	-	-	-	-	-	-	-	-	-
fake-syst (lnN)	-	-	-	-	-	1.3	-	-	-	-
kfactggww (lnN)	-	-	-	-	-	-	-	-	-	-

Table C.2: Pre-fit nuisances (second part).

Appendix D

Validation of the Reweighting Procedure for Generator-Level Samples

In order to produce a smooth map for the exclusion limits, a large number of MC signal samples has to be produced ($\mathcal{O}(10^3)$). Two possible approaches can be followed:

- Produce all the samples with detailed simulation of the interaction of the final state particles with the detector (**FULL-SIM**).
- Produce the additional samples with generation-level information only (**GEN-SIM**).

The first approach would ensure a valid description of the signal, but due to huge number of additional samples, would require a prohibitive amount of time and disk space. The second idea, on the other hand, would ensure a much faster and lighter production. The kinematic properties, at generator level, are used to reweight the relative low number of full-sim MC samples. Assuming the Higgs boson recoils against the DM system, the generator-level Higgs boson p_T is considered as a good variable for such reweighting. In the next section, the technical details of the method, and its validation, are presented.

D.1 Method and Validation

The ratio of the generator-level Higgs boson p_T between the target and the reference samples is used to reweight the fully-simulated sample. The reference mass point to be reweighted in order to obtain the kinematic distributions of the target mass point is chosen in order to minimize the distance in the ($m_{Z'}$,m_A), or in the ($m_{Z'}$,m_χ) plane, with the smaller mass chosen if there are two possibilities. In order to avoid the ratio to present unphysical high values, it is computed by producing histograms of the two distributions (see Figure D.1), and a small constant (0.001) has been added to the content of every bin. The two histograms have been divided bin-by-bin, obtaining a regular distributions. The comparison of the ratio obtained by adding the small constant, and without adding the small constant is presented in Figure D.2, where the regularization effect is observed.
Since adding a small constant to the original p_T distributions can in principle introduce a bias in the ratio, a closure test has been performed. It consists in verifying if it is possible to reproduce the original generator-level Higgs boson p_T by multiplying a reference distribution by the ratio computed as explained:

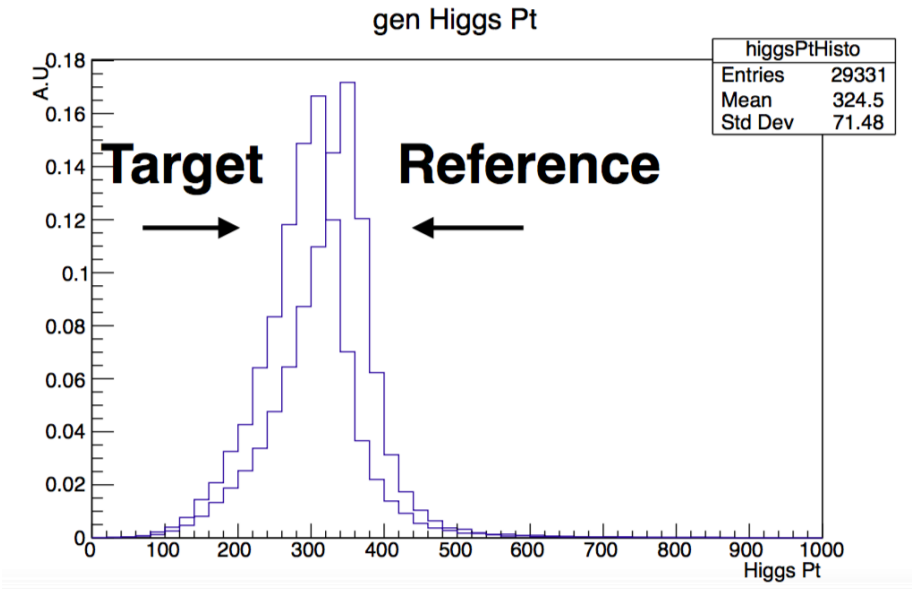
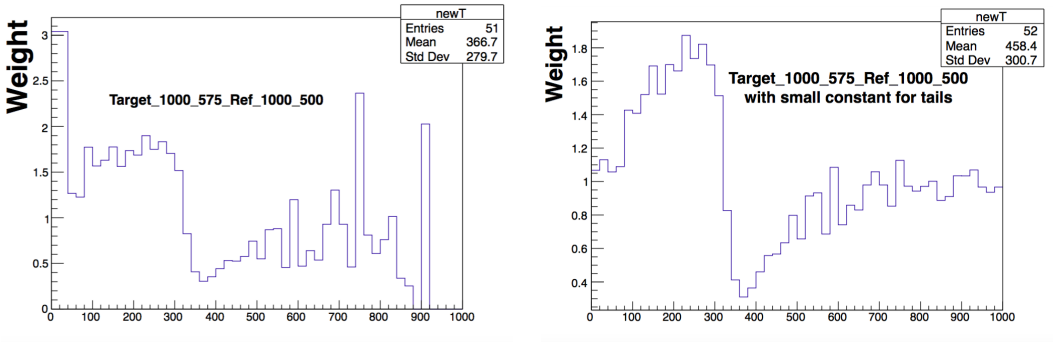


Figure D.1: Distributions of generator-level Higgs boson p_T for two mass points of the Z'-2HDM model ($m_{Z'}$ = 1000 GeV, m_A = 575 GeV vs $m_{Z'}$ = 1000 GeV, m_A = 500 GeV)



(a) Without addining small constant.

(b) Adding small constant.

Figure D.2: Higgs boson p_T ratios, computed without adding a small constant to every bin (left), and by adding a small constant to every bin (right). The regularization effect, in particular in the tail of the distribution, can be clearly observed.

- The generator-level p_T^{Higgs} of one of the full-sim mass points (e.g. $m_{Z'}$ = 1000 GeV, m_A = 500 GeV) is reweighted in order to get the gen-level p_T^{Higgs} of a target gen-sim mass point (e.g. $m_{Z'}$ = 1000 GeV, m_A = 575 GeV);
- This distribution is then divided by the gen-level p_T^{Higgs} of the target mass point (in this case the gen-level p_T^{Higgs} present in the $m_{Z'}$ = 1000 GeV, m_A = 575 GeV mass point);
- If everything works well we expect to get a flat distribution at 1.
    - This means that the method is not putting any bias in the reweighting.

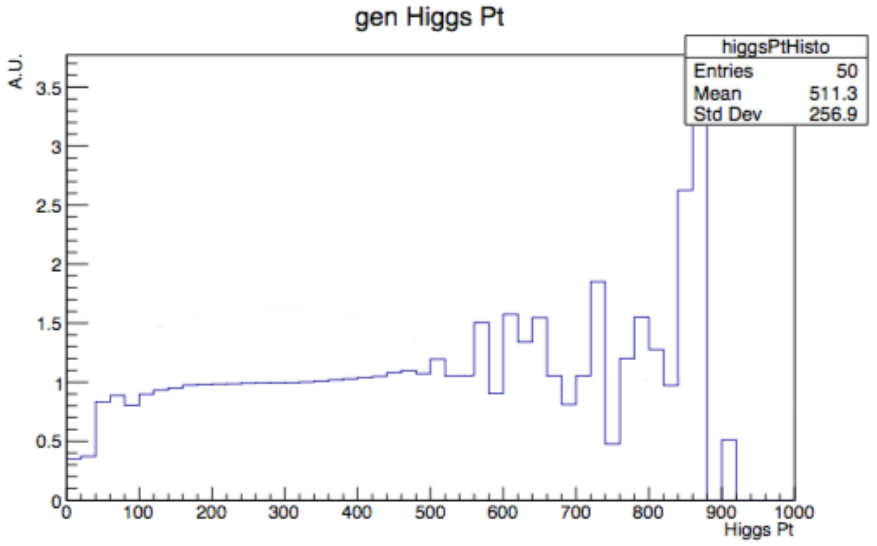
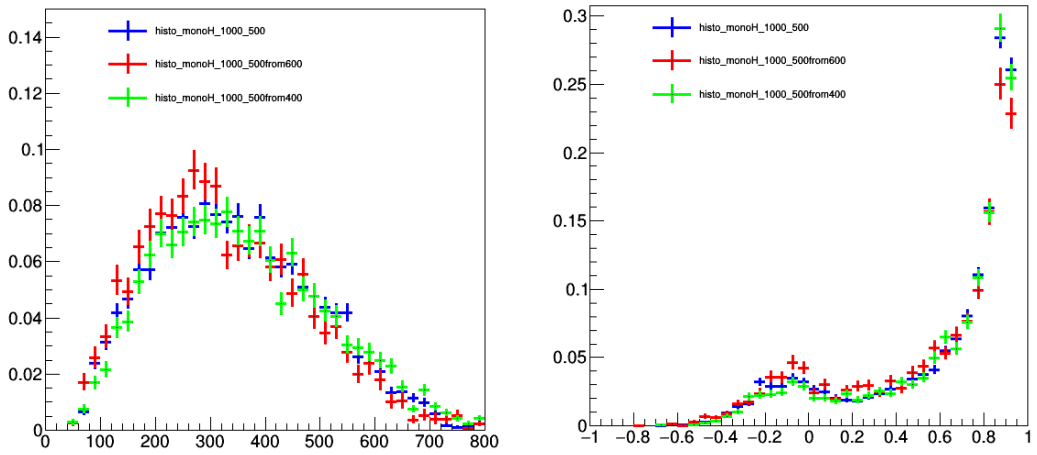


Figure D.3: Ratio between the generator-level Higgs boson p_T of the $m_{Z'}$ = 1000 GeV, m_A = 575 GeV mass point and the p_T^{Higgs} obtained by reweighting the $m_{Z'}$ = 1000 GeV, m_A = 500 GeV mass point. The ratio, in particular in the bulk of the distribution, is smooth and flat around 1, probing that no biases have been introduced by the regularization, obtained by adding 0.001 in each bin of the original distributions.

Once the method to obtain the ratio of the generator-level p_T^{Higgs} has been validated, the assumption that the generator-level p_T^{Higgs} is a good variable to perform the reweighting, and reproduce the kinematics of the mass points produced at generator-level only still has to be verified. This check is performed by comparing some significant kinematic distributions, namely the transverse mass of the Higgs, and the BDT output, obtained with a fully-simulated sample, and obtained by reweighting other fully simulated samples. In particular, for the Z'-2HDM model:

- The target mass point considered: $m_{Z'}$ = 1000 GeV, m_A = 500 GeV;
- The other fully-simulated mass points reweighted to reproduce the kinematic distributions of the target mass points are:
    - $m_{Z'}$ = 1000 GeV, m_A = 400 GeV
    - $m_{Z'}$ = 1000 GeV, m_A = 600 GeV

The results of this check are shown in Figure D.4, where a good agreement between the original distributions, and the distributions obtained with the reweighting is observed. In particular, a slightly better agreement is obtained when the lighter mass point is used for the reweighting.



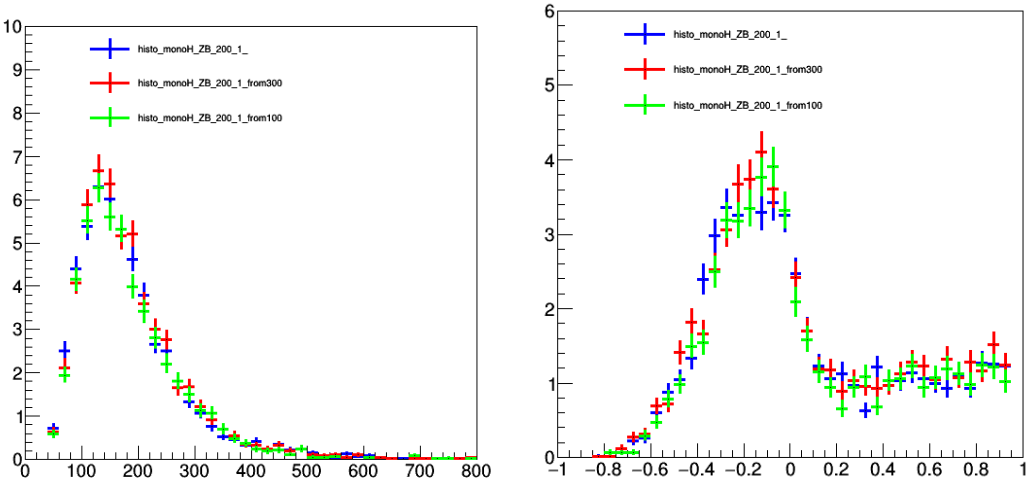
(a) $m_T^{\ell\ell E_T^{miss}}$. (b) BDT.

Figure D.4: $m_T^{\ell\ell E_T^{miss}}$ (left) and BDT output distribution (right) for the Z'-2HDM model. The distributions shown represent the $m_{Z'}$ = 1000 GeV, m_A = 500 GeV, from the full-sim sample (blue), by reweighting the $m_{Z'}$ = 1000 GeV, m_A = 600 GeV full-sim mass point (red), and by reweighting the $m_{Z'}$ = 1000 GeV, m_A = 400 GeV mass point (green).

For the Z'-Baryonic model, on the other hand:

- The target mass point considered: $m_{Z'}$ = 200 GeV, m_χ = 1 GeV;
- The other fully-simulated mass points reweighted to reproduce the kinematic distributions of the target mass points are:
  - $m_{Z'}$ = 300 GeV, m_χ = 1 GeV
  - $m_{Z'}$ = 100 GeV, m_χ = 1 GeV

The results of this check are shown in Figure D.5, where a good agreement between the original distributions, and the distributions obtained with the reweighting is observed. In particular, a slightly better agreement is obtained when the lighter mass point is used for the reweighting. The method can thus be considered as validated for both the models. In particular, since the lighter mass points used for the reweighting show a better agreement with the orginal distributions, if two fully simulated mass points have the same distance in the ($m_{Z'}$,m_A) or ($m_{Z'}$,m_χ) plane from a generator-level mass point to be reweighted, the lightest one is used.



(a) $m_T^{\ell\ell E_T^{miss}}$. (b) BDT.

Figure D.5: $m_T^{\ell\ell E_T^{miss}}$ (left) and BDT output distribution (right) for the Z'-Baryonic model. The distributions shown represent the $m_{Z'} = 200$ GeV, $m_\chi = 1$ GeV, from the full-sim sample (blue), by reweighting the $m_{Z'} = 300$ GeV, $m_\chi = 1$ GeV full-sim mass point (red), and by reweighting the $m_{Z'} = 100$ GeV, $m_\chi = 1$ GeV mass point (green).

List of Figures

List of Tables

Bibliography

[1] A. H. Roberts, M. S. Rots, "Comparison of Rotation Curves of Different Galaxy Types," *Astronomy and Astrophysics*, vol. 26, pp. 483–485, 1973.

[2] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A direct empirical proof of the existence of dark matter," *Astrophys. J.*, vol. 648, pp. L109–L113, 2006.

[3] P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.*, vol. 594, p. A13, 2016.

[4] G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett.*, vol. B716, pp. 1–29, 2012.

[5] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett.*, vol. B716, pp. 30–61, 2012.

[6] L. Carpenter, A. DiFranzo, M. Mulhearn, C. Shimmin, S. Tulin, and D. Whiteson, "Mono-Higgs-boson: A new collider probe of dark matter," *Phys. Rev.*, vol. D89, no. 7, p. 075017, 2014.

[7] D. A. et al., "Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum," 2015.

[8] P. J. E. Peebles and B. Ratra, "The Cosmological constant and dark energy," *Rev. Mod. Phys.*, vol. 75, pp. 559–606, 2003. [,592(2002)].

[9] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints," *Phys. Rept.*, vol. 405, pp. 279–390, 2005.

[10] V. C. Rubin, W. K. Ford, Jr., and N. Thonnard, "Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc," *Astrophys. J.*, vol. 225, pp. L107–L111, 1978.

[11] R. H. Sanders, "The published extended rotation curves of spiral galaxies: confrontation with modified dynamics," *Astrophys. J.*, vol. 473, p. 117, 1996.

[12] C. S. Kochanek, "The Saas Fee Lectures on strong gravitational lensing," in *Proceedings, 33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro: Les Diablerets, Switzerland, April 7-12, 2003*, 2004.

[13] J. E. P. Schneider and E. E. Falco, *Gravitational Lenses.* Springer-Verlag, 1992.

[14] M. Bartelmann and P. Schneider, "Weak gravitational lensing," *Phys. Rept.*, vol. 340, pp. 291–472, 2001.

[15] S. T. S. Institute, "v1.0 data products for epoch 1 of abell 370," 2016.

[16] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker, "Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56," *Astrophys. J.*, vol. 606, pp. 819–824, 2004.

[17] M. Milgrom, "A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis," *Astrophys. J.*, vol. 270, pp. 365–370, 1983.

[18] L. Bergström, "Nonbaryonic dark matter: Observational evidence and detection methods," *Rept. Prog. Phys.*, vol. 63, p. 793, 2000.

[19] P. Tisserand *et al.*, "Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds," *Astron. Astrophys.*, vol. 469, pp. 387–404, 2007.

[20] C. Alcock *et al.*, "The MACHO project: Microlensing results from 5.7 years of LMC observations," *Astrophys. J.*, vol. 542, pp. 281–307, 2000.

[21] P. J. E. Peebles, "The Origin of Galaxies and Clusters of Galaxies.," *Science*, vol. 224, pp. 1385–391, 1984.

[22] J. R. Bond and A. S. Szalay, "The Collisionless Damping of Density Fluctuations in an Expanding Universe," *Astrophys. J.*, vol. 274, pp. 443–468, 1983.

[23] G. Bertone, "Particle dark matter: Observations, Models and Searches," *Cambridge University Press*, 2010.

[24] J. L. Feng, "Dark Matter Candidates from Particle Physics and Methods of Detection," *Ann. Rev. Astron. Astrophys.*, vol. 48, pp. 495–545, 2010.

[25] S. P. Martin, "A Supersymmetry primer," pp. 1–98, 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[26] O. Klein, "Quantentheorie und fünfdimensionale relativitätstheorie," *Z. Phys.*, vol. 37, pp. 895–906, 1926.

[27] K. Hsieh, R. N. Mohapatra, and S. Nasri, "Dark matter in universal extra dimension models: Kaluza-Klein photon and right-handed neutrino admixture," *Phys. Rev.*, vol. D74, p. 066004, 2006.

[28] R. D. Peccei and H. R. Quinn, ""CP" Conservation in the Presence of Pseudoparticles," *Phys. Rev. Lett.*, vol. 38, pp. 1440–1443, Jun 1977.

[29] R. D. Peccei and H. R. Quinn, "Constraints imposed by "CP" conservation in the presence of pseudoparticles," *Phys. Rev. D*, vol. 16, pp. 1791–1797, Sep 1977.

[30] F. Wilczek, "Problem of strong p and t invariance in the presence of instantons," *Phys. Rev. Lett.*, vol. 40, pp. 279–282, Jan 1978.

[31] S. Weinberg, "A new light boson?," *Phys. Rev. Lett.*, vol. 40, pp. 223–226, Jan 1978.

[32] S. Dodelson and L. M. Widrow, "Sterile-neutrinos as dark matter," *Phys. Rev. Lett.*, vol. 72, pp. 17–20, 1994.

[33] T. Saab, "An Introduction to Dark Matter Direct Detection Searches & Techniques," in *The Dark Secrets of the Terascale: Proceedings, TASI 2011, Boulder, Colorado, USA, Jun 6 - Jul 11, 2011*, pp. 711–738, 2013.

[34] J. Liu, X. Chen, and X. Ji, "Current status of direct dark matter detection experiments," *Nature Phys.*, vol. 13, no. 3, pp. 212–216, 2017.

[35] D. S. Akerib *et al.*, "Limits on spin-dependent WIMP-nucleon cross section obtained from the complete LUX exposure," *Phys. Rev. Lett.*, vol. 118, no. 25, p. 251302, 2017.

[36] C. Amole *et al.*, "Dark Matter Search Results from the PICO-60 C_3F_8 Bubble Chamber," *Phys. Rev. Lett.*, vol. 118, no. 25, p. 251301, 2017.

[37] R. Bernabei *et al.*, "Final model independent result of DAMA/LIBRA-phase1," *Eur. Phys. J.*, vol. C73, p. 2648, 2013.

[38] F. Aharonian, "Spectrum and variability of the Galactic Center VHE gamma-ray source HESS J1745-290," 2009.

[39] J. Aleksic *et al.*, "MAGIC Gamma-Ray Telescope Observation of the Perseus Cluster of Galaxies: Implications for Cosmic Rays, Dark Matter and NGC 1275," *Astrophys. J.*, vol. 710, pp. 634–647, 2010.

[40] h. . NASA, title = Fermi Gamma-ray Space Telescope Web Page.

[41] M. Ackermann *et al.*, "Search for Gamma-ray Spectral Lines with the Fermi Large Area Telescope and Dark Matter Implications," *Phys. Rev.*, vol. D88, p. 082002, 2013.

[42] M. Ackermann *et al.*, "The Fermi Galactic Center GeV Excess and Implications for Dark Matter," *Astrophys. J.*, vol. 840, no. 1, p. 43, 2017.

[43] A. Achterberg, M. van Beekveld, S. Caron, G. A. Gómez-Vargas, L. Hendriks, and R. Ruiz de Austri, "Implications of the Fermi-LAT Pass 8 Galactic Center excess on supersymmetric dark matter," *JCAP*, vol. 1712, no. 12, p. 040, 2017.

[44] K. e. a. Choi, "Search for neutrinos from annihilation of captured low-mass dark matter particles in the sun by super-kamiokande," *Phys. Rev. Lett.*, vol. 114, p. 141301, Apr 2015.

[45] M. G. Aartsen *et al.*, "Search for Neutrinos from Dark Matter Self-Annihilations in the center of the Milky Way with 3 years of IceCube/DeepCore," *Eur. Phys. J.*, vol. C77, no. 9, p. 627, 2017.

[46] ATIC collaboration, "ATIC web page." https://cosmicray.umd.edu/atic-home.

[47] O. Adriani *et al.*, "An anomalous positron abundance in cosmic rays with energies 1.5-100 GeV," *Nature*, vol. 458, pp. 607–609, 2009.

[48] M. e. a. Aguilar, "Antiproton flux, antiproton-to-proton flux ratio, and properties of elementary particle fluxes in primary cosmic rays measured with the alpha magnetic spectrometer on the international space station," *Phys. Rev. Lett.*, vol. 117, p. 091103, Aug 2016.

[49] Auger collaboration, "Auger web page." https://www.auger.org/.

[50] T. S., "The First Five Years of the Alpha Magnetic Spectrometer on the International Space Station," *CERN Colloquium*.

[51] L. e. a. Accardo, "High statistics measurement of the positron fraction in primary cosmic rays of 0.5–500 gev with the alpha magnetic spectrometer on the international space station," *Phys. Rev. Lett.*, vol. 113, p. 121101, Sep 2014.

[52] M. Aaboud *et al.*, "Search for new phenomena in dijet events using 37 fb^{-1} of *pp* collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector," *Phys. Rev.*, vol. D96, no. 5, p. 052004, 2017.

[53] M. Aaboud *et al.*, "Search for low-mass dijet resonances using trigger-level jets with the ATLAS detector in *pp* collisions at $\sqrt{s}$ = 13 TeV," *Phys. Rev. Lett.*, vol. 121, no. 8, p. 081801, 2018.

[54] A. M. Sirunyan *et al.*, "Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles," *JHEP*, vol. 08, p. 130, 2018.

[55] A. M. Sirunyan *et al.*, "Search for new physics in dijet angular distributions using proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models," 2018.

[56] M. Aaboud *et al.*, "Search for Dark Matter Produced in Association with a Higgs Boson Decaying to $b\bar{b}$ using 36 fb^{-1} of *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector," *Phys. Rev. Lett.*, vol. 119, no. 18, p. 181804, 2017.

[57] M. Aaboud *et al.*, "Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector," *JHEP*, vol. 01, p. 126, 2018.

[58] M. Aaboud *et al.*, "Search for dark matter at $\sqrt{s} = 13$ TeV in final states containing an energetic photon and large missing transverse momentum with the ATLAS detector," *Eur. Phys. J.*, vol. C77, no. 6, p. 393, 2017.

[59] M. Aaboud *et al.*, "Search for an invisibly decaying Higgs boson or dark matter candidates produced in association with a Z boson in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Lett.*, vol. B776, pp. 318–337, 2018.

[60] "Search for a heavy resonance decaying into a vector boson and a Higgs boson in semileptonic final states at $\sqrt{s} = 13$ TeV," Tech. Rep. CMS-PAS-B2G-17-004, CERN, Geneva, 2018.

[61] "Search for dark matter produced in association with a Higgs boson decaying to $\gamma\gamma$ or $\tau^+\tau^-$ at $\sqrt{s} = 13$ TeV with the CMS detector," Tech. Rep. CMS-PAS-EXO-16-055, CERN, Geneva, 2018.

[62] A. M. Sirunyan *et al.*, "Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV," *Phys. Rev.*, vol. D97, no. 9, p. 092005, 2018.

[63] A. M. Sirunyan *et al.*, "Search for new physics in the monophoton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 10, p. 073, 2017.

[64] "Search for dark matter in final states with a top quark and missing transverse momentum using new hadronic top quark tagging techniques," Tech. Rep. CMS-PAS-EXO-16-051, CERN, Geneva, 2017.

[65] A. M. Sirunyan *et al.*, "Search for new physics in events with a leptonically decaying Z boson and a large transverse momentum imbalance in proton–proton collisions at $\sqrt{s} = 13$ TeV," *Eur. Phys. J.*, vol. C78, no. 4, p. 291, 2018.

[66] S. Banerjee, G. Bélanger, B. Bhattacherjee, F. Boudjema, R. M. Godbole, and S. Mukherjee, "Novel signatures for long-lived particles at the LHC," 2017.

[67] "Search for invisible decays of the Higgs boson produced through vector boson fusion at $\sqrt{s} = 13$ TeV," Tech. Rep. CMS-PAS-HIG-17-023, CERN, Geneva, 2018.

[68] CMS SUSY Public Results. `http://cms-results.web.cern.ch/cms-results/public-results/publications/SUS/13TEV.html`.

[69] ATLAS SUSY Public Results. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/SupersymmetryPublicResultsRetired`.

[70] G. Busoni *et al.*, "Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter," 2016.

[71] ATLAS Collaboration, "Summary plots from the ATLAS Exotic physics group." `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/EXOTICS/`.

[72] CMS Collaboration, "Exotica Summary plots for 13 TeV data." `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryPlotsEXO13TeV`.

[73] "Measurement of antiproton production in pHe collisions at $\sqrt{s_{\rm NN}} = 110$ GeV," Tech. Rep. LHCb-CONF-2017-002. CERN-LHCb-CONF-2017-002, CERN, Geneva, Apr 2017.

[74] M. Thomson, *Modern particle physics*. New York: Cambridge University Press, 2013.

[75] A. Purcell, "Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules," p. 10, Aug 2012.

[76] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.

[77] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.

[78] S. L. Glashow, "Partial Symmetries of Weak Interactions," *Nucl. Phys.*, vol. 22, pp. 579–588, 1961.

[79] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[80] A. Salam and J. C. Ward, "Weak and electromagnetic interactions," *Il Nuovo Cimento*, vol. 11, pp. 568–577, Feb. 1959.

[81] J. Beringer *et al.*, "Review of Particle Physics (RPP)," *Phys. Rev.*, vol. D86, p. 010001, 2012.

[82] G. Aad *et al.*, "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments," *Phys. Rev. Lett.*, vol. 114, p. 191803, 2015.

[83] "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV," Tech. Rep. CMS-PAS-HIG-16-042, CERN, Geneva, 2018.

[84] G. Aad *et al.*, "Observation and measurement of Higgs boson decays to WW* with the ATLAS detector," *Phys. Rev.*, vol. D92, no. 1, p. 012006, 2015.

[85] A. S. et al., "Observation of the higgs boson decay to a pair of τ leptons with the cms detector," *Physics Letters B*, vol. 779, pp. 283 – 316, 2018.

[86] "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector," Tech. Rep. ATLAS-CONF-2018-036, CERN, Geneva, Jul 2018.

[87] M. Aaboud *et al.*, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector," 2018.

[88] A. M. Sirunyan *et al.*, "Observation of $t\bar{t}H$ production," *Phys. Rev. Lett.*, vol. 120, no. 23, p. 231801, 2018.

[89] CMS, "SM Higgs production cross sections at $\sqrt{s} = 13$ TeV." `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV`.

[90] Y. Gershtein *et al.*, "Working Group Report: New Particles, Forces, and Dimensions," in *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013.

[91] T. S. Pettersson and P. Lefèvre, "The Large Hadron Collider: Conceptual design," 1995.

[92] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN Yellow Reports: Monographs, Geneva: CERN, 2004.

[93] G. L. Bayatian *et al.*, "CMS technical design report, volume II: Physics performance," *J. Phys.*, vol. G34, no. 6, pp. 995–1579, 2007.

[94] G. L. Bayatian *et al.*, "CMS Physics," 2006.

[95] *LEP design report*. Geneva: CERN, 1984. Copies shelved as reports in LEP, PS and SPS libraries.

[96] RHIC collaboration, " RHIC web page." https://www.bnl.gov/rhic/.

[97] LHCb collaboration, "LHCb web page." http://lhcb.web.cern.ch/lhcb/.

[98] ALICE collaboration, "ALICE web page." http://alice-collaboration.web.cern.ch/.

[99] T. Mc Cauley, "Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill." CMS Collection., Nov 2016.

[100] T. Kasemets, "Inclusion of Parton Distribution Functions in PYTHIA8," Master's thesis, Lund U., 2010.

[101] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC," *Eur. Phys. J.*, vol. C63, pp. 189–285, 2009.

[102] A. M. Sirunyan *et al.*, "Particle-flow reconstruction and global event description with the CMS detector," *JINST*, vol. 12, no. 10, p. P10003, 2017.

[103] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," Tech. Rep. CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009.

[104] S. Chatrchyan *et al.*, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV," *JINST*, vol. 7, p. P10002, 2012.

[105] R. Kalman, "A new approach to linear filtering and prediction problems," vol. 82D, pp. 35–45, 01 1960.

[106] R. Frühwirth, "Application of kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2, pp. 444 – 450, 1987.

[107] A. A. Anuar, "Electrons and photons at high level trigger in cms for run ii," *Journal of Physics: Conference Series*, vol. 664, no. 8, p. 082001, 2015.

[108] A. Massironi, *Search for the Higgs boson in the WW→ ℓνℓν channel at CMS.* PhD thesis, Milan Bicocca U., 2013-01-15.

[109] and, "On the stopping of fast particles and on the creation of positive electrons," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 146, no. 856, pp. 83–112, 1934.

[110] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP*, vol. 04, p. 063, 2008.

[111] A. M. Sirunyan *et al.*, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *JINST*, vol. 13, no. 05, p. P05011, 2018.

[112] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[113] S. A. et al., "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.

[114] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," *JHEP*, vol. 02, p. 057, 2014.

[115] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory.* Reading, USA: Addison-Wesley, 1995.

[116] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond," *JHEP*, vol. 06, p. 128, 2011.

[117] C. Oleari, "The POWHEG-BOX," *Nucl. Phys. Proc. Suppl.*, vol. 205-206, pp. 36–41, 2010.

[118] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, and P. Torrielli, "Four-lepton production at hadron colliders: aMC@NLO predictions with theoretical uncertainties," *JHEP*, vol. 02, p. 099, 2012.

[119] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An Introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[120] P. Richardson and A. Wilcock, "Monte Carlo Simulation of Hard Radiation in Decays in Beyond the Standard Model Physics in Herwig++," *Eur. Phys. J.*, vol. C74, p. 2713, 2014.

[121] J. Bellm *et al.*, "Herwig++ 2.7 Release Note," 2013.

[122] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.*, vol. 97, pp. 31–145, 1983.

[123] CMS Tag and Probe TWiki. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe`.

[124] A. Berlin, T. Lin, and L.-T. Wang, "Mono-Higgs Detection of Dark Matter at the LHC," *JHEP*, vol. 06, p. 078, 2014.

[125] CMS, "SM Higgs Branching Ratios and Total Decay Widths."

[126] CMS, "Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing." `https://twiki.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns`.

[127] T. C. Collaboration, "Description and performance of track and primary-vertex reconstruction with the cms tracker," *Journal of Instrumentation*, vol. 9, no. 10, p. P10009, 2014.

[128] "Muon pog standard selections for run-ii." https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2.

[129] "Egamma pog standard selections for run-ii." https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2.

[130] JET-MET POG. `https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016`.

[131] BTV POG. `https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods#1a_Event_reweighting_using_scale`.

[132] JET-MET POG, "Met filters recommendations TWiki." `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2`.

[133] "Measurement of the WW cross section pp collisions at sqrt(s)=13 TeV," Tech. Rep. CMS-PAS-SMP-16-006, CERN, Geneva, 2016.

[134] A. M. Sirunyan *et al.*, "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV," *Submitted to: Phys. Lett.*, 2018.

[135] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC," *Nucl. Phys. Proc. Suppl.*, vol. 205-206, pp. 10–15, 2010.

[136] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi, "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD," *Phys. Rev. Lett.*, vol. 113, no. 21, p. 212001, 2014.

[137] P. Meade, H. Ramani, and M. Zeng, "Transverse momentum resummation effects in W^+W^- measurements," *Phys. Rev.*, vol. D90, no. 11, p. 114006, 2014.

[138] P. Jaiswal and T. Okui, "Explanation of the WW excess at the LHC by jet-veto resummation," *Phys. Rev.*, vol. D90, no. 7, p. 073009, 2014.

[139] H. Bakhshian et al., "Computing the contamination from fakes in leptonic final states," *CMS Physics Analysis Note*, vol. CMS-AN-10-261, 2010.

[140] J. Brochero et al., "Data driven estimates of non-prompt leptons background for the higgs boson decaying to ww and ww cross section in the leptonic final state measurements," *CMS Physics Analysis Note*, vol. CMS-AN-13-184, 2013.

[141] ATLAS and L. H. C. G. CMS Collaborations, "Procedure for the LHC Higgs boson search combination in Summer 2011," CMS Note ATL-PHYS-PUB 2011-11, CMS-NOTE 2011-005, ATLAS and CMS Collaborations, 2011.

[142] S. Chatrchyan *et al.*, "Combined results of searches for the standard model higgs boson in pp collisions at $\sqrt{s} = 7$ *tev*," *Phys. Lett. B*, vol. 710, p. 26, 2012.

[143] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J.*, vol. C71, p. 1554, 2011. [Erratum: Eur. Phys. J.C73,2501(2013)].

[144] S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B*, vol. 716, p. 30, 2012.

[145] CMS, "Luminosity POG TWiki." `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[146] "Ecal dpg electron scale." `https://indico.cern.ch/event/402608/contribution/1/attachments/1206871/1758949/DPG_reReco_corrections.pdf`.

[147] N. T. et al., "Higgs boson decaying to ww in the leptonic final state at 13 tev," CMS Note 2016/182, CERN, 2016.

[148] JET-MET POG, "JET-MET POG TWiki." `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetMET`.

[149] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II," *J. Phys.*, vol. G43, p. 023001, 2016.

[150] D. E. De Florian Sabaris, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher, and R. Tanaka, "Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector," Sep 2016.

[151] S. Chatrchyan *et al.*, "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states," *JHEP*, vol. 01, p. 096, 2014.

[152] "Nnlo+nnll top-quark-pair cross sections." `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO`.

[153] "Nlo single-top channel cross sections." `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec`.

[154] F. Caola, K. Melnikov, R. Rotsch, and L. Tancredi, "QCD corrections to W^+W^- production through gluon fusion," *Phys. Lett.*, vol. B754, pp. 275–280, 2016.